



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Avi ASHKENAZI, *et al.*

Application Serial No. 09/993,604

Filed: November 14, 2001

For: **SECRETED AND TRANSMEMBRANE**
POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME

) Examiner: Landsman, Robert S.

) Art Unit: 1647

) Confirmation No: 1800

) Attorney's Docket No. 39780-2730 P1C25

) Customer No. 77845

EXPRESS MAIL LABEL NO.: EB 985 660 805 US

DATE MAILED: March 18, 2008

ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES
APPELLANTS' BRIEF

MAIL STOP APPEAL BRIEF - PATENTS

Commissioner for Patents

P.O. Box 1450

Alexandria, Virginia 22313-1450

Dear Sir:

This Appeal Brief, filed in connection with the above captioned patent application, is responsive to the Final Office Action mailed on September 18, 2007. A Notice of Appeal was filed herein on December 18, 2007. This Appeal Brief is being filed with a one month extension of time. Appellants hereby appeal to the Board of Patent Appeals and Interferences from the final rejection in this case.

The following constitutes the Appellants' Brief on Appeal.

03/20/2008 HVUONG1 00000046 071700 09993604

01 FC:1402 510.00 DA

03/20/2008 HVUONG1 00000046 071700 09993604

02 FC:1251 120.00 DA

I. REAL PARTY IN INTEREST

The real party in interest is Genentech, Inc., South San Francisco, California, by an assignment of the parent application, U.S. Patent Application Serial No. 09/941,992 recorded November 16, 2001, at Reel 012176 and Frame 0450.

II. RELATED APPEALS AND INTERFERENCES

The claims pending in the current application are directed to a polypeptide referred to herein as "PRO1281". There exist two related patent applications, (1) U.S. Serial No. 09/989,726, filed November 19, 2001 now Patent No. 7,018,811, issued 03-28-2006 (containing claims directed to nucleic acids encoding PRO1281 polypeptides), and (2) U.S. Serial No. 09/997,542, filed November 15, 2001 (containing claims directed to antibodies to PRO1281 polypeptides). The related U.S. Serial No. 09/997,542 application is also under final rejection by the same Examiner and based upon the same outstanding rejections, an appeal is being pursued independently and concurrently herewith.

III. STATUS OF CLAIMS

Claims 119-126 and 129-131 are in this application.

Claims 1-118 and 127-128 have been canceled.

Claims 119-126 and 129-131 stand rejected and Appellants appeal the rejection of these claims.

IV. STATUS OF AMENDMENTS

A summary of the prosecution history for this case is as follows:

Previously, in response to a Final Office Action mailed on August 30, 2006, a Notice of Appeal was filed December 1, 2006 and an Appeal Brief was filed June 30, 2007. This was followed by another Final Office action wherein the finality of the previous Office Action was withdrawn by the Examiner. A Notice of Appeal was filed on December 18, 2007 in this case.

No claim amendments have been submitted after the last final rejection of September 18, 2007. All previous amendments have been entered. A copy of the rejected claims in the present Appeal is provided in Section IX.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The invention claimed in the present application is related to an isolated polypeptide comprising the amino acid sequence of the polypeptide of SEQ ID NO: 326, referred to in the present application as "PRO1281." The PRO1281 gene was shown for the first time in the present application to be significantly amplified in human colon cancers as compared to normal, non-cancerous human tissue controls (Example 170). This feature is specifically recited in Claim 124, and carried by all claims dependent from Claim 124. In addition, the invention also claims the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal-peptide; or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129 (Claims 124-126 and 129). The invention is further directed to polypeptides having at least 80% to 99% amino acid sequence identity to the amino acid sequence of the polypeptide of SEQ ID NO: 326; the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide; or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129, wherein the nucleic acid encoding said polypeptide is amplified in colon tumor (Claims 119-123). The invention is further directed to a chimeric polypeptide comprising one of the above polypeptides fused to a heterologous polypeptide (Claim 130), and to a chimeric polypeptide wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin (Claim 131). PRO polypeptide variants having at least about 95-99% amino acid sequence identity with a full length PRO polypeptide sequence, or a PRO polypeptide sequence lacking the signal peptide are generally described in the specification at, for example, page 305, line 23 onwards, and percent amino acid sequence identity determination is generally described at least at, for example, pages 306-308, line 14 onwards. The preparation of chimeric PRO polypeptides (Claims 130 and 131), including those wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin, is set forth in the specification at page 374, lines 24 to page 375, line 9. Examples 140-143 and page 376, line 12 onwards describe the expression of PRO polypeptides in various host cells, including *E. coli*, mammalian cells, yeast and Baculovirus-infected insect cells.

The amino acid sequence of the native "PRO1281" polypeptide and the nucleic acid sequence encoding this polypeptide (referred to in the present application as "DNA59820-1549") are shown in the present specification as SEQ ID NOs: 326 and 325, respectively, and in Figures 151 and 150, described on pages 295, lines 3-5. The full-length PRO1281 polypeptide having the amino acid sequence of SEQ ID NO: 326 is described in the specification at, for example, on page 18 and pages 140-142 and the isolation of cDNA clones encoding PRO1281 of SEQ ID NO: 326 is described in Example 64, page 454-455 of the specification.

Finally, Example 170, in the specification at page 539, line 19, to page 555, line 5, sets forth a 'Gene Amplification assay' which shows that the PRO1281 gene is amplified in the genome of certain human colon cancers (see Table 9A, page 550). The profiles of various primary colon tumors used for screening the PRO polypeptide compounds of the invention in the gene amplification assay are summarized on Table 8, page 546 of the specification.

VI. GROUND OF REJECTION TO BE REVIEWED ON APPEAL

1. (a) Whether Claims 119-126 and 129-131 satisfy the utility/ enablement requirement under 35 U.S.C. §§101/112, first paragraph.
(b) Whether Claims 119-123 and 129-131 satisfy the enablement requirement under 35 U.S.C. §112, first paragraph.
2. Whether Claims 119-123 and 129-131 satisfy the written description requirement under 35 U.S.C. §112, first paragraph.
3. Whether Claims 119-126 and 129-131 are anticipated under 35 U.S.C. §102(b) by Baker (WO99/63088- dated December 1999).

VII. ARGUMENTS

Summary of the Arguments

Issue 1(a): Utility/ Enablement

Appellants rely upon the gene amplification data of the PRO1281 gene for patentable utility of the PRO1281 polypeptides. This data is clearly disclosed in the instant specification in Example 170 which discloses that the gene encoding PRO1281 showed significant amplification,

ranging from **2.099 fold to 2.219-fold** in different colon primary tumors. Example 170 in the instant specification discloses that "(a)mplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as colon, lung, breast and other cancers and diagnostic determination of the presence of those cancers" (emphasis added). Therefore, such a gene is useful as a marker for the diagnosis of colon cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

Appellants have submitted, with their Response filed July 22, 2005, the Declaration by Dr. Audrey Goddard, which explains that a gene identified as being amplified at least 2-fold by the disclosed gene amplification assay in a tumor sample relative to a normal sample is useful as a marker, for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

Appellants further submitted a number of references as well as Declarations by experts in the field of oncology, during prosecution (for example, Declarations by Dr. Ashkenazi and Dr. Goddard), that collectively show that, in general, if a gene is amplified in cancer, it is more likely than not that the corresponding mRNA and the encoded protein will be expressed at an elevated level. For instance, the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* (made of record in Appellants' Response filed June 4, 2004) collectively teach that in general, gene amplification increases mRNA expression. As the Examiner no longer takes a position that mRNA levels are not predictive of polypeptide levels, the evidence presented by Appellants support that gene amplification correlates with the increased protein expression. Based on the **2.099 fold to 2.219-fold** amplification of PRO1281 in colon primary tumors, one of ordinary skill would find it credible that the claimed PRO1281 polypeptides would have utility as markers for the diagnosis of colon tumors, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

Further, Appellants submit that one of ordinary skill in the art would know how to make and use the recited polypeptides for the diagnosis of colon cancer without any undue experimentation, based on the detailed teachings in the specification.

Accordingly, this enablement rejection under 35 U.S.C. §§101 and 112, first paragraph should be withdrawn.

Issue 1(b): Scope of Enablement

Appellants note that the claims recite structural features, namely, 80-99% sequence identity to SEQ ID NO: 326, which are common to the genus. The genus of claimed polypeptides is further defined by having a specific functional activity for the encoding nucleic acids, namely, that the encoding nucleic acid is amplified in colon tumors. The specification provides detailed guidance as to how to identify the recited variants of SEQ ID NO: 326, including methods for determining percent identity between two amino acid sequences, as well as listings of exemplary and preferred sequence substitutions, as well as detailed protocols for determining whether a gene encoding a variant PRO1281 protein is amplified in colon tumor. Thus one of skill in the art could easily make and use a variant PRO1281 for the diagnosis of colon cancer without any undue experimentation.

Accordingly, this enablement rejection under 35 U.S.C. §112, first paragraph should be withdrawn.

Issue 2: Written Description

Appellants note that the claims recite structural features, namely, 80-99% sequence identity to SEQ ID NO: 326, which are common to the genus. The genus of claimed polypeptides is further defined by having a specific functional activity for the encoding nucleic acids, namely, that the encoding nucleic acid is amplified in colon tumors. The specification provides detailed guidance as to how to identify the recited variants of SEQ ID NO: 326, including methods for determining percent identity between two amino acid sequences, as well as listings of exemplary and preferred sequence substitutions, as well as detailed protocols for determining whether a gene encoding a variant PRO1281 protein is amplified in colon tumor. Thus one of skill in the art could easily identify whether a variant PRO1281 sequence falls within the parameters of the claimed invention.

Accordingly, a description of the claimed genus has been achieved by the recitation of both structural and functional characteristics. The Written Description rejection under 35 U.S.C. §112, first paragraph should be withdrawn.

Issue 3: Anticipation by Baker *et al.* (December 1999)

Claims 119-126 and 129-131 remain rejected under 35 U.S.C. §102(b) as being anticipated by Baker *et al.* (WO 99/63088, December 1999).

The instant application claims PRO1281 polypeptides. For the same reasons discussed under Issue 2 regarding utility, Appellants submit that U.S. provisional application 60/141037 also has utility, based on the gene amplification assay disclosed in that application. Appellants have made a proper priority claim to U.S. provisional application 60/141037, filed June 23, 1999 and should be accorded to the priority date of June 23, 1999 for the instant application. Therefore, Baker *et al.* is not prior art.

These arguments are all discussed in further detail below under the appropriate headings.

Response to Rejections

ISSUE 1(a). Claims 119-126 and 129-131 are supported by a credible, specific and substantial asserted utility, and thus meet the utility requirement of 35 U.S.C. §§101/112, first paragraph

The sole basis for the Examiner's rejection of Claims 119-126 and 129-131 under this section is that the data presented in Example 170 of the present specification is allegedly insufficient to establish a patentable utility under 35 U.S.C. §101 for the presently claimed PRO1281 polypeptides.

Claims 119-126 and 129-131 stand further rejected under 35 U.S.C. §112, first paragraph, allegedly "since the claimed invention is not supported by either a specific and substantial asserted utility or a well established utility for the reasons set forth above, one skilled in the art clearly would not know how to use the claimed invention."

Appellants strongly disagree and, therefore, respectfully traverse the rejection.

A. The Legal Standard For Utility Under 35 U.S.C. §101

According to 35 U.S.C. §101:

Whoever invents or discovers any new and *useful* process, machine, manufacture, or composition of matter, or any new and *useful* improvement thereof, may obtain

a patent therefor, subject to the conditions and requirements of this title.
(Emphasis added).

In interpreting the utility requirement, in *Brenner v. Manson*,¹ the Supreme Court held that the *quid pro quo* contemplated by the U.S. Constitution between the public interest and the interest of the inventors required that a patent Appellant disclose a "substantial utility" for his or her invention, *i.e.*, a utility "where specific benefit exists in currently available form."² The Court concluded that "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion. A patent system must be related to the world of commerce rather than the realm of philosophy."³

Later, in *Nelson v. Bowler*,⁴ the C.C.P.A. acknowledged that tests evidencing pharmacological activity of a compound may establish practical utility, even though they may not establish a specific therapeutic use. The Court held that "since it is crucial to provide researchers with an incentive to disclose pharmaceutical activities in as many compounds as possible, we conclude adequate proof of any such activity constitutes a showing of practical utility."⁵

In *Cross v. Iizuka*,⁶ the C.A.F.C. reaffirmed *Nelson*, and added that *in vitro* results might be sufficient to support practical utility, explaining that "*in vitro* testing, in general, is relatively less complex, less time consuming, and less expensive than *in vivo* testing. Moreover, *in vitro* results with the particular pharmacological activity are generally predictive of *in vivo* test results,

¹ *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. (BNA) 689 (1966).

² *Id.* at 534, 148 U.S.P.Q. (BNA) at 695.

³ *Id.* at 536, 148 U.S.P.Q. (BNA) at 696.

⁴ *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. (BNA) 881 (C.C.P.A. 1980).

⁵ *Id.* at 856, 206 U.S.P.Q. (BNA) at 883.

⁶ *Cross v. Iizuka*, 753 F.2d 1047, 224 U.S.P.Q. (BNA) 739 (Fed. Cir. 1985).

i.e. there is a reasonable correlation there between."⁷ The Court perceived, "No insurmountable difficulty" in finding that, under appropriate circumstances, "*in vitro* testing, may establish a practical utility."⁸

The case law has also clearly established that Appellants' statements of utility are usually sufficient, unless such statement of utility is unbelievable on its face.⁹ The PTO has the initial burden to prove that Appellants' claims of usefulness are not believable on their face.¹⁰ In general, an Appellant's assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. §101, "unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope."^{11, 12}

Compliance with 35 U.S.C. §101 is a question of fact.¹³ The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the totality of the evidence under consideration.¹⁴ Thus, to overcome the presumption of truth that an assertion of utility by the Appellant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the Appellant. The issue will then be decided on the totality of evidence.

⁷ *Id.* at 1050, 224 U.S.P.Q. (BNA) at 747.

⁸ *Id.*

⁹ *In re Gazave*, 379 F.2d 973, 154 U.S.P.Q. (BNA) 92 (C.C.P.A. 1967).

¹⁰ *Ibid.*

¹¹ *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. (BNA) 288, 297 (C.C.P.A. 1974).

¹² See also *In re Jolles*, 628 F.2d 1322, 206 U.S.P.Q. 885 (C.C.P.A. 1980); *In re Irons*, 340 F.2d 974, 144 U.S.P.Q. 351 (1965); *In re Sichert*, 566 F.2d 1154, 1159, 196 U.S.P.Q. 209, 212-13 (C.C.P.A. 1977).

¹³ *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. (BNA) 592, 596 (Fed. Cir. 1983) *cert. denied*, 469 US 835 (1984).

¹⁴ *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d (BNA) 1443, 1444 (Fed. Cir. 1992).

The well established case law is clearly reflected in the Utility Examination Guidelines (“Utility Guidelines”),¹⁵ which acknowledge that an invention complies with the utility requirement of 35 U.S.C. §101, if it has at least one asserted “specific, substantial, and credible utility” or a “well-established utility.” Under the Utility Guidelines, a utility is “specific” when it is particular to the subject matter claimed. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic without also identifying the conditions that are to be diagnosed.

In explaining the “substantial utility” standard, M.P.E.P. §2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase “immediate benefit to the public” or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be “currently available” to the public in order to satisfy the utility requirement. “Rather, any reasonable use that an applicant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a ‘substantial’ utility.”¹⁶ Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement,¹⁷ gives the following instruction to patent examiners: “If the Appellant has asserted that the claimed invention is useful for any particular practical purpose . . . and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility.”

B. Proper Application of the Legal Standard

Appellants respectfully submit that the data presented in Example 170 starting on page 539 of the specification and the cumulative evidence of record, indeed support a “specific, substantial and credible” asserted utility for the presently claimed invention.

Patentable utility for the PRO1281 polypeptides is based upon the gene amplification data for the gene encoding the PRO1281 polypeptide. Example 170 describes the results

¹⁵ 66 Fed. Reg. 1092 (2001).

¹⁶ M.P.E.P. §2107.01.

¹⁷ M.P.E.P. §2107 II(B)(1).

obtained using a very well-known and routinely employed polymerase chain reaction (PCR)-based assay, the TaqManTM PCR assay, also referred to herein as the gene amplification assay. This assay allows one to quantitatively measure the level of gene amplification in a given sample, say, a tumor extract, or a cell line. It was well known in the art at the time the invention was made that gene amplification is an essential mechanism for oncogene activation. Appellants isolated genomic DNA from a variety of primary cancers and cancer cell lines that are listed in Table 9 (pages 539 onwards of the specification), including primary colon cancers of the type and stage indicated in Table 8 (page 546). The tumor samples were tested in triplicates with TaqmanTM primers and with internal controls, beta-actin and GADPH in order to quantitatively compare DNA levels between samples (page 548, lines 33-34). As a negative control, DNA was isolated from the cells of ten normal healthy individuals, which was pooled and used as a control (page 539, lines 27-29) and also, no-template controls (page 548, lines 33-34). The results of TaqManTM PCR are reported in Δ Ct units, as explained in the passage on page 539, lines 37-39. One unit corresponds to one PCR cycle or approximately a 2-fold amplification, relative to control, two units correspond to 4-fold, 3 units to 8-fold amplification and so on. Using this PCR-based assay, Appellants showed that the gene encoding for PRO1281 was amplified, that is, it showed approximately 1.07-1.15 Δ Ct units which corresponds to $2^{1.07}$ - $2^{1.15}$ - fold amplification or **2.099 fold to 2.219-fold** in different colon primary tumors.

In support of their showing that these gene amplification values are significant, Appellants submitted, in the RCE filed July 22, 2005, a Declaration by Dr. Audrey Goddard. Appellants particularly draw the Board's attention to page 3 of the Goddard Declaration which clearly states that:

It is further my considered scientific opinion that an at least **2-fold increase** in gene copy number in a tumor tissue sample relative to a normal (*i.e.*, non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is **useful as a marker for the diagnosis of cancer**, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

(Emphasis added).

Accordingly, the **2.099 fold to 2.219-fold** in different colon primary tumors would be considered significant and credible by one skilled in the art, based upon the facts disclosed in the Goddard Declaration.

The Examiner has asserted that only two out of the fourteen colon cancer samples tested positive for the claimed polypeptide, thus, if a sample was taken from an individual with colon cancer for diagnosis, it is more likely than not that this assay would yield a false negative result. (Page 3 of the Office Action mailed September 18, 2007)

Appellants point out that, as any skilled artisan in the field of oncology would easily appreciate, not all tumor markers are generally associated with every tumor, or even, with most tumors. In fact, some tumor markers are useful for identifying rare malignancies. That is, the association of the tumor marker with a particular type of tumor lesion may be rare, or, the occurrence of that particular kind of tumor lesion itself may be rare. In either event, even these rare tumor markers, which may not give a positive hit with most common tumors, have great value in tumor diagnosis, and consequently, in tumor prognosis. The skilled artisan would know that such tumor markers are very useful for better classification of tumors. Therefore, whether the PRO1281 gene is amplified in most of the colon tumors is not relevant to its identification as a tumor marker, or its patentable utility. Rather, whether the amplification data for PRO1281 is significant is what lends support to its usefulness as a tumor marker. It was well known in the art at the time of filing of the application that gene amplification, which occurs in most solid tumors like colon cancers, is generally associated with poor prognosis. Therefore, the PRO1281 gene becomes an important diagnostic marker to identify such malignant colon cancers, even if the malignancy associated with PRO1281 molecule is a rare occurrence and the present specification clearly discloses enough evidence such that, one skilled in the art would know that PRO1281 is useful as a valuable diagnostic marker for identifying certain types of colon cancers.

The Examiner has also alleged, based on Sen *et al.*, that the observed gene amplification was due to aneuploidy. The Examiner has also asserted it is not clear whether PRO1281 is

amplified in cancerous colon tissue more than in damaged colon tissue. (Page 3 of the Office Action mailed September 18, 2007)

Appellants respectfully disagree and submit that their gene amplification data was not due to aneuploidy. Appellants had submitted the Ashkenazi Declaration to show that "detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy." Regarding Sen, Appellants agree that while aneuploidy can be a feature of damaged tissue as well, besides cancerous or pre-cancerous tissue, and may not invariably lead to cancer, Sen *et al.* in fact support the Appellants' position that PRO1281 is still useful in diagnosing pre-cancerous lesions or cancer itself. For instance, the art in colon cancer at the time of filing of the instant application clearly described that "epithelial tumors develop through a multistep process driven by genetic instability" in damaged colon lesions which may eventually lead to colon cancer. Many articles published around June 23, 1999 (the effective filing date of this application) studied such damaged or premalignant lesions and suggested that identification of such pre-cancerous lesions were very important in preventive diagnosis and treatment of colon cancer. Based on the well-known art, Appellants submit that there is utility in identifying genetic biomarkers in epithelial tissues at cancer risk.

A prima facie case of lack of utility has not been established

The Examiner has asserted that the gene amplification data discussed above is not sufficient to provide utility for the PRO1281 or the claimed antibodies that bind it. In particular, the Examiner has asserted that it does not necessarily follow that an increase in gene copy number results in increased gene expression and increased protein expression, such that the polypeptide would be useful diagnostically. The Examiner has cited a reference by Pennica *et al.* and Konopka *et al.* to support his position (Page 3 of the Office Action mailed September 18, 2007).

As a preliminary matter, Appellants respectfully submit that it is not a legal requirement to establish that gene amplification "necessarily" results in increased expression at the mRNA and polypeptide levels or that polypeptide levels can be "accurately predicted." As discussed above, the evidentiary standard to be used throughout *ex parte* examination of a patent application is a preponderance of the totality of the evidence under consideration. Accordingly,

Appellants submit that in order to overcome the presumption of truth that an assertion of utility by the applicant enjoys, the Examiner must establish that **it is more likely than not** that one of ordinary skill in the art would doubt the truth of the statement of utility. Therefore, it is not legally required that there be a “necessary” correlation between the data presented and the claimed subject matter. The law requires only that one skilled in the art should accept that such a correlation is **more likely than not to exist**. Appellants respectfully submit that when the proper evidentiary standard is applied, a correlation must be acknowledged.

Appellants submit that Pennica *et al.* does not show a lack of correlation between gene (DNA) amplification and mRNA levels. According to the quoted statement from Pennica *et al.*, "WISP-1 gene amplification in human colon tumors showed a correlation between DNA amplification and over-expression, whereas overexpression of WISP-3 RNA was seen in the absence of DNA amplification. In contrast, WISP-2 DNA was amplified in colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with expression in normal colonic mucosa from the same patient." From this, the Examiner correctly concludes that increased copy number does not *necessarily* result in increased polypeptide expression. The standard, however, is not absolute certainty. The fact that in the case of a specific class of closely related molecules there seemed to be no correlation with gene amplification and the level of mRNA/protein expression, does not establish that it is more likely than not, in general, that such correlation does not exist. The Examiner has not shown whether the lack or correlation observed for the family of WISP polypeptides is typical, or is merely a discrepancy, an exception to the rule of correlation. Indeed, the working hypothesis among those skilled in the art is that, if a gene is amplified in cancer, the encoded protein is likely to be expressed at an elevated level. In fact, as noted even in Pennica *et al.*, “[a]n analysis of *WISP-1* gene amplification and expression in human colon tumors *showed a correlation between DNA amplification and over-expression . . .*” (Pennica *et al.*, page14722, left column, first full paragraph, emphasis added).

Accordingly, Appellants respectfully submit that Pennica *et al.* teaches nothing conclusive regarding the absence of correlation between amplification of a gene and over-expression of the encoded WISP polypeptide. More importantly, the teaching of Pennica *et al.* is

specific to *WISP* genes. Pennica *et al.* has no teaching whatsoever about the correlation of gene amplification and protein expression in general.

Similarly, in Konopka *et al.*, Appellants submit that the Examiner has generalized a very specific result disclosed by Konopka *et al.* to cover all genes. Konopka *et al.* actually state that “[p]rotein expression is not related to amplification of the *abl* gene but to variation in the level of *bcr-abl* mRNA produced from a single Ph¹ template.” (See Konopka *et al.*., Abstract, emphasis added). The paper does not teach anything whatsoever about the correlation of protein expression and gene amplification in general, and provides no basis for the generalization that apparently underlies the present rejection. The statement of Konopka *et al.* that “[p]rotein expression is not related to amplification of the *abl* gene . . . ” is not sufficient to establish a *prima facie* case of lack of utility. Therefore, the combined teachings of Pennica *et al.* and Konopka *et al.* are not directed towards genes in general but to a single gene or genes within a single family and thus, their teachings cannot support a general conclusion regarding correlation between gene amplification and mRNA or protein levels.

Regarding the Godbout reference, Appellants respectfully submit that it was never claimed that PRO1281 is similar in any way to the DDX1 gene of Godbout *et al.*, they never claimed PRO1281 was an RNA helicase or that it confers selective advantage to cell survival; on the other hand, the Godbout reference was submitted to show good correlation between protein levels based upon genomic DNA amplification, which the Examiner clearly agrees with. Moreover, selective advantage to cell survival is not the only mechanism by which genes impact cancer. Structure/function data, which the Examiner requests, is not a requirement for the utility requirement. Hence this rejection is improper.

Appellants respectfully further point out that Li *et al.* acknowledge that their results differed from those obtained by Hyman *et al.* and Pollack *et al.* (of record), who found a substantially higher level of correlation between gene amplification and increased gene expression. The authors note that “[t]his discordance may reflect methodologic differences between studies or biological differences between breast cancer and lung adenocarcinoma” (page 2629, col. 1). For instance, as explained in the Supplemental Information accompanying the Li article, genes were considered to be amplified if they had a copy number ratio of at least 1.40. In

the case of PRO1281, as discussed in previously filed responses and in the Goddard Declaration (of record), an appropriate threshold for considering gene amplification to be significant is a copy number of at least 2.0 (which is a higher threshold). The PRO1281 gene showed significant amplification of **2.099 fold to 2.219-fold** in different colon primary tumors, and thus fully meets this standard. It is not surprising that, in the Li *et al.* reference, by using a lower threshold of 1.4 for considering gene amplification, a higher number of genes not showing corresponding increases in mRNA expression were found. Nevertheless, the results of Li *et al.* do not conclusively disprove that a gene with a substantially higher level of gene amplification, such as PRO1281, would be expected to show a corresponding increase in transcript expression.

It is "more likely than not" for amplified genes to have increased mRNA

On the contrary, Appellants submit that Example 170 of the specification further discloses that, "(a)mplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as lung, colon, breast and other cancers and diagnostic determination of the presence of those cancers" (Emphasis added). Besides, Appellants have submitted ample evidence to show that, in general, if a gene is amplified in cancer, it is "more likely than not" that the encoded protein will also be expressed at an elevated level.

For instance, Appellants presented the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* (made of record in Appellants' Response filed November 4, 2004), who collectively teach that in general, for most genes, DNA amplification increases mRNA expression. The results presented by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* are based upon wide ranging analyses of a large number of tumor associated genes. Orntoft *et al.* studied transcript levels of 5600 genes in malignant bladder cancers, many of which were linked to the gain or loss of chromosomal material, and found that in general (18 of 23 cases) chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Hyman *et al.* compared DNA copy numbers and mRNA expression of over 12,000 genes in breast cancer tumors and cell lines, and found that there was evidence of a prominent global influence of copy number changes on gene expression levels. In Pollack *et al.*, the authors

profiled DNA copy number alteration across 6,691 mapped human genes in 44 predominantly advanced primary breast tumors and 10 breast cancer cell lines, and found that on average, a 2-fold change in DNA copy number was associated with a corresponding 1.5-fold change in mRNA levels. In summary, the evidence supports the Appellants' position that gene amplification is more likely than not predictive of increased mRNA and polypeptide levels. Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip® arrays in 2004. Clearly, the research community believe that the information obtained from these chips is useful (*i.e.*, that it is more likely than not that the results are informative of protein levels).

Thus, the Examiner appears to disregard the ample evidence provided in the above referenced articles based on misinterpretations of their teachings. Appellants submit that in fact, these articles lend significant support that for an amplified gene, it is more likely than not that the protein will also be overexpressed and would be viewed as reasonable and credible by one of ordinary skill in the art. The "more likely than not" standard is a much lower standard than a "necessary" correlation or "accurate" prediction, and is clearly met in the claimed invention. Moreover, the Examiner has not cited any evidence or advanced any arguments as to why Appellants' statement of overexpression of protein would not be credible. Accordingly, this point is believed to be moot.

The Examiner has stated that Appellants have not indicated whether PRO1281 is in a gene cluster region of a chromosome. (Page 7 of the Office Action mailed September 18, 2007). Appellants fail to see how this is relevant to the analysis. Orntoft *et al.* did not limit their findings to only those regions of amplified gene clusters. Further, as discussed below, Hyman *et al.* and Pollack *et al.* did gene-by-gene analysis across all chromosomes.

Appellants respectfully submit that the Examiner has mischaracterized the methods used by Hyman *et al.* and Pollack *et al.* in their analysis. These papers did not use traditional CGH analysis to identify amplified genes. In Hyman *et al.*, 13,824 cDNA clones were placed on glass slides in a microarray and genomic DNA from breast cancer cell lines and normal human WBCs was hybridized to the cDNA sequences. For expression analysis, RNA from tumor cell lines was

hybridized on the same microarrays. The 13,824 arrayed cDNA clones were analyzed for gene expression and gene copy number in 14 breast cancer cell lines. Hyman *et al.* state, "The results illustrate a considerable influence of copy number on gene expression patterns." For example, Hyman *et al.* teach that "[u]p to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (*i.e.*, belonged to the global upper 7% of expression ratios) compared with only 6% for genes with normal copy number." (See page 6242, column 1). Further, Hyman *et al.* state that "[t]he cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome." (See page 6242, column 2). Therefore, the analysis performed by Hyman *et al.* was on a gene-by gene basis, and clearly shows that "it is more likely than not" that a gene which is amplified in tumor cells will have increased gene expression.

In Pollack *et al.*, DNA copy number alteration across 6,691 mapped human genes in 44 predominantly advanced primary breast tumors and 10 breast cancer cell lines was profiled. Pollack *et al.* further state, "Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells." (See Abstract). "Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4, and representing 91 different genes), 62% (representing 54 different genes; ...) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4)." (See page 12966, column 1). Therefore, the analysis performed by Pollack *et al.* was also on a gene-by gene basis, and clearly shows that "it is more likely than not" that a gene which is amplified in tumor cells will have increased gene expression.

The Examiner has further asserted that "none of the three papers reported that the research was relevant to identifying probes that can be used as cancer diagnostics" (Page 8 of the Office Action mailed September 18, 2007). Appellants respectfully point out that Hyman *et al.* conducted additional studies of one of the genes found to be amplified, HOXB7, and found "a clinical association between HOXB7 amplification and poor patient prognosis." (Page 6244, col.1 to col.2). Thus the results of Hyman *et al.* confirm that genes which are amplified in

tumors have prognostic utility. The Board's attention is also respectfully directed to the final paragraph of Pollack *et al.*, wherein the authors conclude that "a substantial portion of the phenotypic uniqueness (and, by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number." (Page 12698, col. 2). Accordingly, Pollack *et al.* confirm that genes that are amplified in at least one type of tumor are useful as markers for that type of tumor, and for prognostic uses directed to that type of tumor.

Even if a *prima facie* case of lack of utility has been established, it should be withdrawn on consideration of the totality of evidence

Even if one assumes *arguendo* that it is more likely than not that there is no correlation between gene amplification and increased mRNA/protein expression, which Appellants submit is **not** true, a polypeptide encoded by a gene that is amplified in cancer would **still** have a specific, substantial, and credible utility. In support, Appellants respectfully draw the Board's attention to page 2 of the Declaration of Dr. Avi Ashkenazi (submitted with the Response filed June 4, 2004) which explains that,

even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

Appellants thus submit that simultaneous testing of gene amplification and gene product over-expression enables more accurate tumor classification, even if the gene-product, the protein, is not over-expressed. This leads to better determination of a suitable therapy. Further, as explained in Dr. Ashkenazi's Declaration, absence of over-expression of the protein itself is crucial information for the practicing clinician. If a gene is amplified in a tumor, but the

corresponding gene product is not over-expressed, the clinician will decide not to treat a patient with agents that target that gene product. This not only saves money, but also has the benefit that the patient can avoid exposure to the side effects associated with such agents.

This utility is further supported by the teachings of the article by Hanna and Mornin. (Pathology Associates Medical Laboratories, August (1999), submitted with the Response filed June 4, 2004). The article teaches that the HER-2/neu gene has been shown to be amplified and/or over-expressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinomas. Further, the article teaches that diagnosis of breast cancer includes testing both the amplification of the HER-2/neu gene (by FISH) as well as the over-expression of the HER-2/neu gene product (by IHC). Even when the protein is not over-expressed, the assay relying on both tests leads to a more accurate classification of the cancer and a more effective treatment of it.

The Examiner has asserted that "Hanna et al. supports the rejection, in that Hanna et al. show that gene amplification does not reliably correlate with protein over-expression, and thus the level of polypeptide expression must be tested empirically." (Page 9 of the Office Action mailed September 18, 2007). Appellants respectfully point out that the Examiner appears to have misread Hanna *et al.* Hanna *et al.* clearly state that gene amplification (as measured by FISH) and polypeptide expression (as measured by immunohistochemistry, IHC) are well correlated ("in general, FISH and IHC results correlate well" (Hanna *et al.* p. 1, col. 2)). It is only a subset of tumors which show discordant results. Thus Hanna *et al.* support Appellants' position that it is more likely than not that gene amplification correlates with increased polypeptide expression.

Appellants have clearly shown that the gene encoding the PRO1281 polypeptide is amplified in at least two colon tumors. Therefore, the PRO1281 gene, similar to the HER-2/neu gene disclosed in Hanna *et al.*, is a tumor associated gene. Furthermore, as discussed above, in the majority of amplified genes, the teachings in the art overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Therefore, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO1281 gene, that the PRO1281 polypeptide is concomitantly overexpressed.

The Examiner appears to view the testing described in the Ashkenazi Declaration and the Hanna article as experiments involving further characterization of the PRO1281 polypeptide itself. In fact, such testing is for the purpose of characterizing not the PRO1281 polypeptide, but the tumors in which the gene encoding PRO1281 is amplified. The PRO1281 polypeptide and the claimed antibodies which bind it are therefore useful in tumor categorization, the results of which become an important tool in the hands of a physician enabling the selection of a treatment modality that holds the most promise for the successful treatment of a patient.

ISSUE 2(b). Claims 119-123 and 129-131 satisfy the enablement requirement of 35 U.S.C.

§112, First Paragraph

Claims 119-123 and 129-131 stand rejected under 35 U.S.C. §112, first paragraph, as allegedly lacking enablement because “the standard for enablement is ‘make and use’ not ‘make and test’” (page 5 of the Final Office action mailed August 30, 2006). Appellants respectfully traverse this rejection.

Appellants note that the claims recite structural features, namely, 80-99% sequence identity to SEQ ID NO: 326, which are common to the genus. The genus of claimed polypeptides is further defined by having a specific functional activity for the encoding nucleic acids, namely, that the encoding nucleic acid is amplified in colon tumors. As is discussed in greater detail below under Issue 3, the specification provides detailed guidance as to how to identify the recited variants of SEQ ID NO: 326, including **methods for determining percent identity between two amino acid sequences**, as well as listings of exemplary and preferred sequence substitutions, as well as detailed protocols for determining whether a gene encoding a variant PRO1281 protein is amplified in colon tumor. Only those PRO1281 variants that have the “function” of their gene being amplified in colon tumors are claimed herein. Thus one of skill in the art could easily identify whether a variant PRO1281 sequence falls within the parameters of the claimed invention, and would not view it as an “invitation to test” the invention.

Accordingly, a description of the claimed genus has been achieved by the recitation of both structural and functional characteristics and one of ordinary skill in the art would

understand how to make and use the recited polypeptide for the diagnosis of colon cancer without any undue experimentation.

Accordingly, this enablement rejection under 35 U.S.C. §112, first paragraph is misplaced and should be withdrawn.

ISSUE 2: Claims 119-123 and 129-131 satisfy the written description requirement of 35 U.S.C. §112, First Paragraph

Claims 119-123 and 129-131 stand rejected under 35 U.S.C. §112, first paragraph, as allegedly containing "subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention." In particular, the Examiner has asserted that "(Applicant) does not provide specific information about SEQ ID NO: 326" (Page 6, last paragraph of the Final Office Action mailed August 30, 2006).

Appellants respectfully disagree.

A. The Legal Test for Written Description

The well-established test for sufficiency of support under the written description requirement of 35 U.S.C. §112, first paragraph is "whether the disclosure of the application as originally filed reasonably conveys to the artisan that the inventor had possession at that time of the later claimed subject matter, rather than the presence or absence of literal support in the specification for the claim language."^{18, 19} The adequacy of written description support is a factual issue and is to be determined on a case-by-case basis.²⁰ The factual determination in a

¹⁸ *In re Kaslow*, 707 F.2d 1366, 1374, 212 U.S.P.Q. 1089, 1096 (Fed. Cir. 1983).

¹⁹ *See also Vas-Cath, Inc. v. Mahurkar*, 935 F.2d at 1563, 19 U.S.P.Q.2d at 1116 (Fed. Cir. 1991).

²⁰ *See e.g., Vas-Cath*, 935 F.2d at 1563; 19 U.S.P.Q.2d at 1116.

written description analysis depends on the nature of the invention and the amount of knowledge imparted to those skilled in the art by the disclosure.^{21, 22}

In *Environmental Designs, Ltd. v. Union Oil Co.*,²³ the Federal Circuit held, "Factors that may be considered in determining level of ordinary skill in the art include: (1) the educational level of the inventor; (2) type of problems encountered in the art; (3) prior art solutions to those problems; (4) rapidity with which innovations are made; (5) sophistication of the technology; and (6) educational level of active workers in the field."²⁴ Further, the "hypothetical 'person having ordinary skill in the art' to which the claimed subject matter pertains would, of necessity have the capability of understanding the scientific and engineering principles applicable to the pertinent art."^{25, 26}

B. The Disclosure Provides Sufficient Written Description for the Claimed Invention

Appellants respectfully submit that the instant specification evidences the actual reduction to practice of the amino acid sequence of SEQ ID NO: 326. Appellants also submit that the specification provides ample written support for determining percent sequence identity between two amino acid sequences (See pages 306-308, line 14 onwards). In fact, the specification teaches specific parameters to be associated with the term "percent identity" as applied to the present invention. The specification further provides detailed guidance as to changes that may be made to a PRO polypeptide without adversely affecting its activity (page 372, line 36 to page 373, line 17). This guidance includes a listing of exemplary and

²¹ *Union Oil v. Atlantic Richfield Co.*, 208 F.2d 989, 996 (Fed. Cir. 2000).

²² *See also* M.P.E.P. §2163 II(A).

²³ 713 F.2d 693, 696, 218 U.S.P.Q. 865, 868 (Fed. Cir. 1983), *cert. denied*, 464 U.S. 1043 (1984).

²⁴ *See also* M.P.E.P. §2141.03.

²⁵ *Ex parte Hiyamizu*, 10 U.S.P.Q.2d 1393, 1394 (Bd. Pat. App. & Inter. 1988) (emphasis added).

²⁶ *See also* M.P.E.P. §2141.03.

preferred substitutions for each of the twenty naturally occurring amino acids (Table 6, page 372). Accordingly, one of skill in the art could identify whether the variant PRO1281 sequence falls within the parameters of the claimed invention. Once such an amino acid sequence was identified, the specification sets forth methods for making the amino acid sequences (see page 376, line 9) and methods of preparing the PRO polypeptides (see Examples 140-143). Applicants submit that one skilled in the art would know that Applicants possessed these variants based on this disclosure.

Currently pending Claims 119-123 and 129-131 recite the functional recitation that the nucleic acid encoding the claimed polypeptides are amplified in colon tumors. Appellants further submit that the specification provides ample written support for detecting and quantifying amplification of such nucleic acids in several tumors and/or cell lines as described in Example 170. Example 170 of the present application provides step-by-step guidelines and protocols for the gene amplification assay. By following this disclosure, one skilled in the art would know that it is easy to test whether a gene encoding a variant PRO1281 protein is amplified in colon tumors by the methods set forth in Example 170.

More recently, in *Enzo Biochem., Inc. v. Genprobe, Inc.* 296 F.3d 1316 (Fed. Cir. 2002), the court adopted the standard that "the written description requirement can be met by 'showing that the invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics, . . . i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics." *Id.* at 1324. While the invention in *Enzo* was still a DNA, the holding has been treated as being applicable to proteins as well. Indeed, the court adopted the standard from the USPTO's Written Description Examination Guidelines, which apply to both proteins and nucleic acids.

Accordingly, current applicable case law holds that biological sequences are not adequately described solely by a description of their desired functional activities. The instant claims meet the standard set by the *Enzo* court in that the claimed sequences are defined not only by functional properties, but also by structural limitations. It is well established that a combination of functional and structural features may suffice to describe a claimed genus. "An

applicant may also show that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics which provide evidence that applicant was in possession of the claimed invention, i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics."²⁷ Thus, the genus of polypeptides with at least 80-99% sequence identity to SEQ ID NO: 326, which possess the functional property of having a nucleic acid which is amplified in colon tumor would meet the requirement of 35 U.S.C. §112, first paragraph, as providing adequate written description. Accordingly, one skilled in the art would have known that Appellants had knowledge and possessed the claimed polypeptides with 80-99% sequence identity to SEQ ID NO: 326 whose encoding nucleic acids were amplified in colon tumors. The recited property of amplification of the encoding gene adds to the characterization of the claimed polypeptide sequences in a manner that one of skill in the art could readily assess and understand.

As discussed above, Appellants have recited structural features, namely, 80-99% sequence identity to SEQ ID NO: 326, which are common to the genus. Appellants have also provided guidance as to how to make the recited variants of SEQ ID NO: 326, **including listings of exemplary and preferred sequence substitutions**. The genus of claimed polypeptides is further defined by having a specific functional activity for the encoding nucleic acids. Accordingly, a description of the claimed genus has been achieved.

Accordingly, Appellants respectfully request reconsideration and reversal of the written description rejection of Claims 119-123 and 129-131 under 35 U.S.C. §112, first paragraph.

ISSUE 3. Claims 119-126, 129-131 and 135-138 are not anticipated by Baker *et al.*

Claims 119-126 and 129-131 remain rejected under 35 U.S.C. §102(b) for being anticipated by Baker (dated December 1999).

The instant application claims **PRO1281 polypeptides**. For the same reasons discussed in detail under Issue 2, Appellants maintain that they rely on the gene amplification assay for

²⁷ M.P.E.P. §2163 II(A)(3)(a).

patentable utility which was first disclosed in U.S. Provisional Application 60/141037, filed June 23, 1999, priority to which has been properly claimed in the instant application. Hence, the present application is entitled to at least the priority date of **June 23, 1999**.

Therefore, Baker *et al.* is not prior art. Accordingly, this rejection under 35 U.S.C. §102(b) should be withdrawn.

CONCLUSION

For the reasons given above, Appellants submit that present specification clearly describes, details and provides a patentable utility for the claimed invention. Moreover, it is respectfully submitted that based upon this disclosed patentable utility, the present specification clearly teaches "how to use" the presently claimed polypeptide. As such, Appellants respectfully request reconsideration and reversal of the outstanding rejection of Claims 119-126 and 129-131.

The Commissioner is authorized to charge any fees which may be required, including extension fees, or credit any overpayment to Deposit Account No. **07-1700** (referencing Attorney's Docket No. **39780-2730 P1C25**).

Respectfully submitted,

Date: March 18, 2008

By: 
PANPAN GAO, Registration No. 43,626

GOODWIN PROCTER LLP
181 lytton Avenue
Palo Alto, California 94301
Telephone: (650) 752-3100
Facsimile: (650) 853-1038

VIII. CLAIMS APPENDIX

119. An isolated native sequence polypeptide having at least 80% amino acid sequence identity to:

- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 326;
- (b) the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide;
- (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129, wherein, the nucleic acid encoding said polypeptide is amplified in colon tumors.

120. The isolated native sequence polypeptide of Claim 39 having at least 85% amino acid sequence identity to:

- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 326;
- (b) the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide;
- (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129, wherein, the nucleic acid encoding said polypeptide is amplified in colon tumors.

121. The isolated native sequence polypeptide of Claim 39 having at least 90% amino acid sequence identity to:

- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 326;
- (b) the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide;
- (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129, wherein, the nucleic acid encoding said polypeptide is amplified in colon tumors.

122. The isolated native sequence polypeptide of Claim 39 having at least 95% amino acid sequence identity to:

- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 326;
- (b) the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide;
- (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129, wherein, the nucleic acid encoding said polypeptide is amplified in colon tumors.

123. The isolated native sequence polypeptide of Claim 39 having at least 99% amino acid sequence identity to:

- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 326;
- (b) the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide,
- (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129, wherein, the nucleic acid encoding said polypeptide is amplified in colon tumors.

124. An isolated polypeptide comprising:

- (a) the amino acid sequence of the polypeptide of SEQ ID NO: 326;
- (b) the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide,
- (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129, wherein, the nucleic acid encoding said polypeptide is amplified in colon tumors.

125. The isolated polypeptide of Claim 124 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 326.

126. The isolated polypeptide of Claim 124 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 326, lacking its associated signal peptide.

129. The isolated polypeptide of Claim 124 comprising the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203129.

130. A chimeric polypeptide comprising a polypeptide according to Claim 124 fused to a heterologous polypeptide.

131. The chimeric polypeptide of Claim 130, wherein said heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin.

IX. EVIDENCE APPENDIX

1. Declaration of Audrey Goddard, Ph.D. under 35 C.F.R. §1.132, with attached Exhibits A-G:

- A. Curriculum Vitae of Audrey D. Goddard, Ph.D.
- B. Higuchi, R. *et al.*, "Simultaneous amplification and detection of specific DNA sequences," *Biotechnology* 10:413-417 (1992).
- C. Livak, K.J., *et al.*, "Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization," *PCR Methods Appl.* 4:357-362 (1995).
- D. Heid, C.A. *et al.*, "Real time quantitative PCR," *Genome Res.* 6:986-994 (1996).
- E. Pennica, D. *et al.*, "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors," *Proc. Natl. Acad. Sci. USA* 95:14717-14722 (1998).
- F. Pitti, R.M. *et al.*, "Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer," *Nature* 396:699-703 (1998).
- G. Bieche, I. *et al.*, "Novel approach to quantitative polymerase chain reaction using real-time detection: Application to the detection of gene amplification in breast cancer," *Int. J. Cancer* 78:661-666 (1998).
2. Declaration of Avi Ashkenazi, Ph.D. under 35 C.F.R. §1.132, with attached Exhibit A (Curriculum Vitae).
3. Orntoft, T.F., *et al.* Molecular & Cellular Proteomics – 1:37-45 (2002).
4. Hyman, E., *et al.*, "Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer," *Cancer Research* 62:6240-6245 (2002).
5. Pollack, J.R., *et al.*, "Microarray Analysis Reveals a Major Direct Role of DNA Copy Number Alteration in the Transcriptional Program of Human Breast Tumors," *Proc. Natl. Acad. Sci. USA* 99:12963-12968 (2002).
6. Hanna *et al.*, "HER-2/neu Breast Cancer Predictive Testing," Pathology Associates Medical Laboratories (1999).
7. Pennica, *et al.* "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors" *Proc. Natl. Acad. Sci. USA Vol. 95, pp. 14717-14722, December 1998*
8. Konopka *et al.*, "Variable Expression of the Translocated c-abl oncogene in Philadelphia-chromosome-positive B-lymphoid cell lines from chronic myelogenous leukemia patients" *Proc. Natl. Acad. Sci. USA* 83: 4049-52, (1986).
9. Sen *et al.* , *Current Opinion in Oncology*, 12: 82-88, (2000). (cited June 2, 2003)
10. Godbout, R., *et al.*, *J. Biol. Chem.* - 273(33):21161-8 (1998).
11. Li *et al.*, 2006, *Oncogene* 25: 2628-2635.

Item 1 was submitted with Appellants' Response filed July 22, 2005, and was considered by the Examiner as indicated in the Final Office Action mailed August 11, 2005.

Items 2-6 were submitted with Appellants' Response filed June 4, 2004, and were considered by the Examiner as indicated in the Final Office Action mailed July 26, 2004.

Items 7-8 were made of record by the Examiner as indicated in the Office Action mailed July 26, 2006.

Item 9 was made of record by the Examiner in the Final Office Action mailed August 11, 2005.

Item 10 was submitted with Appellant's Preliminary Amendment filed June 29, 2006 and acknowledged by the Examiner on August 30, 2006.

Item 11 was made of record by the Examiner in the Final Office Action mailed September 18, 2007.

X. RELATED PROCEEDINGS APPENDIX

None- no decision rendered by a Court or the Board in any related proceedings identified above.

SV 2320874 v1
18/03/2008 1:07 PM (39780.2730)

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Ashkenazi et al.	Group Art Unit: 1647
Serial No.: 09/903,925	Examiner: Fozia Hamid
Filed: July 11, 2001	CERTIFICATE OF MAILING I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Assistant Commissioner of Patents, Washington, D.C. 20231 on _____ Date _____
For: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS	

DECLARATION OF AUDREY D. GODDARD, Ph.D UNDER 37 C.F.R. § 1.132

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

I, Audrey D. Goddard, Ph.D. do hereby declare and say as follows:

1. I am a Senior Clinical Scientist at the Experimental Medicine/BioOncology, Medical Affairs Department of Genentech, Inc., South San Francisco, California 94080.
2. Between 1993 and 2001, I headed the DNA Sequencing Laboratory at the Molecular Biology Department of Genentech, Inc. During this time, my responsibilities included the identification and characterization of genes contributing to the oncogenic process, and determination of the chromosomal localization of novel genes.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

Serial No.: *
Filed: *

4. I am familiar with a variety of techniques known in the art for detecting and quantifying the amplification of oncogenes in cancer, including the quantitative TaqMan PCR (i.e., "gene amplification") assay described in the above captioned patent application.

5. The TaqMan PCR assay is described, for example, in the following scientific publications: Higuchi *et al.*, Biotechnology 10:413-417 (1992) (Exhibit B); Livak *et al.*, PCR Methods Appl., 4:357-362 (1995) (Exhibit C) and Heid *et al.*, Genome Res. 6:986-994 (1996) (Exhibit D). Briefly, the assay is based on the principle that successful PCR yields a fluorescent signal due to Taq DNA polymerase-mediated exonuclease digestion of a fluorescently labeled oligonucleotide that is homologous to a sequence between two PCR primers. The extent of digestion depends directly on the amount of PCR, and can be quantified accurately by measuring the increment in fluorescence that results from decreased energy transfer. This is an extremely sensitive technique, which allows detection in the exponential phase of the PCR reaction and, as a result, leads to accurate determination of gene copy number.

6. The quantitative fluorescent TaqMan PCR assay has been extensively and successfully used to characterize genes involved in cancer development and progression. Amplification of protooncogenes has been studied in a variety of human tumors, and is widely considered as having etiological, diagnostic and prognostic significance. This use of the quantitative TaqMan PCR assay is exemplified by the following scientific publications: Pennica *et al.*, Proc. Natl. Acad. Sci. USA 95(25):14717-14722 (1998) (Exhibit E); Pitti *et al.*, Nature 396(6712):699-703 (1998) (Exhibit F) and Bieche *et al.*, Int. J. Cancer 78:661-666 (1998) (Exhibit G), the first two of which I am co-author. In particular, Pennica *et al.* have used the quantitative TaqMan PCR assay to study relative gene amplification of WISP and c-myc in various cell lines, colorectal tumors and normal mucosa. Pitti *et al.* studied the genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, using the quantitative TaqMan PCR assay. Bieche *et al.* used the assay to study gene amplification in breast cancer.

Serial No.: *

Filed: *

7. It is my personal experience that the quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that an at least 2-fold increase in gene copy number in a tumor tissue sample relative to a normal (i.e., non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

8. I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Jan. 16, 2003

Date

Audrey D. Goddard

Audrey D. Goddard, Ph.D.

AUDREY D. GODDARD, Ph.D.

Genentech, Inc.
1 DNA Way
South San Francisco, CA, 94080
650.225.6429
goddarda@gene.com

110 Congo St.
San Francisco, CA, 94131
415.841.9154
415.819.2247 (mobile)
agoddard@pacbell.net

PROFESSIONAL EXPERIENCE

Genentech, Inc.
South San Francisco, CA

1993-present

2001 - present Senior Clinical Scientist
Experimental Medicine / BioOncology, Medical Affairs

Responsibilities:

- *Companion diagnostic oncology products*
- *Acquisition of clinical samples from Genentech's clinical trials for translational research*
- *Translational research using clinical specimen and data for drug development and diagnostics*
- *Member of Development Science Review Committee, Diagnostic Oversight Team, 21 CFR Part 11 Subteam*

Interests:

- *Ethical and legal implications of experiments with clinical specimens and data*
- *Application of pharmacogenomics in clinical trials*

1998 - 2001 Senior Scientist

Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities:

- *Management of a laboratory of up to nineteen –including postdoctoral fellow, associate scientist, senior research associate and research assistants/associate levels*
- *Management of a \$750K budget*
- *DNA sequencing core facility supporting a 350+ person research facility.*
- *DNA sequencing for high throughput gene discovery, - ESTs, cDNAs, and constructs*
- *Genomic sequence analysis and gene identification*
- *DNA sequence and primary protein analysis*

Research:

- *Chromosomal localization of novel genes*
- *Identification and characterization of genes contributing to the oncogenic process*
- *Identification and characterization of genes contributing to inflammatory diseases*
- *Design and development of schemes for high throughput genomic DNA sequence analysis*
- *Candidate gene prediction and evaluation*

1993 - 1998 Scientist

Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities

- *DNA sequencing core facility supporting a 350+ person research facility*
- *Assumed responsibility for a pre-existing team of five technicians and expanded the group into fifteen, introducing a level of middle management and additional areas of research*
- *Participated in the development of the basic plan for high throughput secreted protein discovery program – sequencing strategies, data analysis and tracking, database design*
- *High throughput EST and cDNA sequencing for new gene identification.*
- *Design and implementation of analysis tools required for high throughput gene identification.*
- *Chromosomal localization of genes encoding novel secreted proteins.*

Research:

- *Genomic sequence scanning for new gene discovery.*
- *Development of signal peptide selection methods.*
- *Evaluation of candidate disease genes.*
- *Growth hormone receptor gene SNPs in children with Idiopathic short stature*

**Imperial Cancer Research Fund
London, UK with Dr. Ellen Solomon**

1989-1992

6/89 –12/92 Postdoctoral Fellow

- Cloning and characterization of the genes fused at the acute promyelocytic leukemia translocation breakpoints on chromosomes 17 and 15.
- Prepared a successfully funded European Union multi-center grant application

**McMaster University
Hamilton, Ontario, Canada with Dr. G. D. Sweeney**

1983

5/83 – 8/83: NSERC Summer Student

- *In vitro* metabolism of β -naphthoflavone in C57Bl/6J and DBA mice

EDUCATION

Ph.D.

"Phenotypic and genotypic effects of mutations in the human retinoblastoma gene."

Supervisor: Dr. R. A. Phillips

University of Toronto
Toronto, Ontario, Canada.
Department of Medical
Biophysics.

1989

Honours B.Sc

"The *in vitro* metabolism of the cytochrome P-448 inducer β -naphthoflavone in C57BL/6J mice."

Supervisor: Dr. G. D. Sweeney

McMaster University,
Hamilton, Ontario, Canada.
Department of Biochemistry

1983

ACADEMIC AWARDS

Imperial Cancer Research Fund Postdoctoral Fellowship	1989-1992
Medical Research Council Studentship	1983-1988
NSERC Undergraduate Summer Research Award	1983
Society of Chemical Industry Merit Award (Hons. Biochem.)	1983
Dr. Harry Lyman Hooker Scholarship	1981-1983
J.L.W. Gill Scholarship	1981-1982
Business and Professional Women's Club Scholarship	1980-1981
Wyerhauser Foundation Scholarship	1979-1980

INVITED PRESENTATIONS

Genentech's gene discovery pipeline: High throughput identification, cloning and characterization of novel genes. Functional Genomics: From Genome to Function, Litchfield Park, AZ, USA. October 2000

High throughput identification, cloning and characterization of novel genes. G2K:Back to Science, Advances in Genome Biology and Technology I. Marco Island, FL, USA. February 2000

Quality control in DNA Sequencing: The use of Phred and Phrap. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. April 1999

High throughput secreted protein identification and cloning. Tenth International Genome Sequencing and Analysis Conference, Miami, FL, USA. September 1998

The evolution of DNA sequencing: The Genentech perspective. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. May 1998

Partial Growth Hormone Insensitivity: The role of GH-receptor mutations in Idiopathic Short Stature. Tenth Annual National Cooperative Growth Study Investigators Meeting, San Francisco, CA, USA. October, 1996

Growth hormone (GH) receptor defects are present in selected children with non-GH-deficient short stature: A molecular basis for partial GH-insensitivity. 76th Annual Meeting of The Endocrine Society, Anaheim, CA, USA. June 1994

A previously uncharacterized gene, myl, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. XV International Association for Comparative Research on Leukemia and Related Disease, Padua, Italy. October 1991

PATENTS

Goddard A, Godowski PJ, Gurney AL. NL2 Tie ligand homologue polypeptide. Patent Number: 6,455,496. Date of Patent: Sept. 24, 2002.

Goddard A, Godowski PJ and Gurney AL. NL3 Tie ligand homologue nucleic acids. Patent Number: 6,426,218. Date of Patent: July 30, 2002.

Godowski P, Gurney A, Hillan KJ, Botstein D, **Goddard A**, Roy M, Ferrara N, Tumas D, Schwall R. NL4 Tie ligand homologue nucleic acid. Patent Number: 6,413,770. Date of Patent: July 2, 2002.

Ashkenazi A, Fong S, **Goddard A**, Gurney AL, Napier MA, Tumas D, Wood WI. Nucleic acid encoding A-33 related antigen poly peptides. Patent Number: 6,410,708. Date of Patent: Jun. 25, 2002.

Botstein DA, Cohen RL, **Goddard AD**, Gurney AL, Hillan KJ, Lawrence DA, Levine AJ, Pennica D, Roy MA and Wood WI. WISP polypeptides and nucleic acids encoding same. Patent Number: 6,387,657. Date of Patent: May 14, 2002.

Goddard A, Godowski PJ and Gurney AL. Tie ligands. Patent Number: 6,372,491. Date of Patent: April 16, 2002.

Godowski PJ, Gurney AL, **Goddard A** and Hillan K. TIE ligand homologue antibody. Patent Number: 6,350,450. Date of Patent: Feb. 26, 2002.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Tie receptor tyrosine kinase ligand homologues. Patent Number: 6,348,351. Date of Patent: Feb. 19, 2002.

Goddard A, Godowski PJ and Gurney AL. Ligand homologues. Patent Number: 6,348,350. Date of Patent: Feb. 19, 2002.

Attie KM, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 6,207,640. Date of Patent: March 27, 2001.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Nucleic acids encoding NL-3. Patent Number: 6,074,873. Date of Patent: June 13, 2000

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,824,642. Date of Patent: October 20, 1998

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,646,113. Date of Patent: July 8, 1997

Multiple additional provisional applications filed

PUBLICATIONS

- Seshasayee D, Dowd P, Gu Q, Erickson S, **Goddard AD**. Comparative sequence analysis of the *HER2* locus in mouse and man. Manuscript in preparation.
- Abuzzahab MJ, **Goddard A**, Grigorescu F, Lautier C, Smith RJ and Chernausk SD. Human IGF-1 receptor mutations resulting in pre- and post-natal growth retardation. Manuscript in preparation.
- Aggarwal S, Xie, M-H, Foster J, Frantz G, Stinson J, Corpuz RT, Simmons L, Hillan K, Yansura DG, Vandlen RL, **Goddard AD** and Gurney AL. FHFR, a novel receptor for the fibroblast growth factors. Manuscript submitted.
- Adams SH, Chui C, Schilbach SL, Yu XX, **Goddard AD**, Grimaldi JC, Lee J, Dowd P, Colman S., Lewin DA. (2001) BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene, and assessment of a potential link to obesity. *Biochemical Journal* **360**: 135-142.
- Lee J, Ho WH, Maruoka M, Corpuz RT, Baldwin DT, Foster JS, **Goddard AD**, Yansura DG, Vandlen RL, Wood WI, Gurney AL. (2001) IL-17E, a novel proinflammatory ligand for the IL-17 receptor homolog IL-17Rh1. *Journal of Biological Chemistry* **276**(2): 1660-1664.
- Xie M-H, Aggarwal S, Ho W-H, Foster J, Zhang Z, Stinson J, Wood WI, **Goddard AD** and Gurney AL. (2000) Interleukin (IL)-22, a novel human cytokine that signals through the interferon-receptor related proteins CRF2-4 and IL-22R. *Journal of Biological Chemistry* **275**: 31335-31339.
- Weiss GA, Watanabe CK, Zhong A, **Goddard A** and Sidhu SS. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA* **97**: 8950-8954.
- Guo S, Yamaguchi Y, Schilbach S, Wada T., Lee J, **Goddard A**, French D, Handa H, Rosenthal A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.
- Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, **Goddard A**, de Vos AM, Gao WQ, Dixit VM. (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**: 523-527.
- Sehl PD, Tai JTN, Hillan KJ, Brown LA, **Goddard A**, Yang R, Jin H and Lowe DG. (2000) Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* **101**: 1990-1999.
- Guo S, Brush J, Teraoka H, **Goddard A**, Wilson SW, Mullins MC and Rosenthal A. (1999) Development of noradrenergic neurons in the zebrafish hindbrain requires BMP, FGF8, and the homeodomain protein soulless/Phox2A. *Neuron* **24**: 555-566.
- Stone D, Murone, M, Luoh, S, Ye W, Armanini P, Gurney A, Phillips HS, Brush, J, **Goddard A**, de Sauvage FJ and Rosenthal A. (1999) Characterization of the human suppressor of fused; a negative regulator of the zinc-finger transcription factor Gli. *J. Cell Sci.* **112**: 4437-4448.
- Xie M-H, Holcomb I, Deuel B, Dowd P, Huang A, Vagts A, Foster J, Liang J, Brush J, Gu Q, Hillan K, **Goddard A** and Gurney, A.L. (1999) FGF-19, a novel fibroblast growth factor with unique specificity for FGFR4. *Cytokine* **11**: 729-735.

- Yan M, Lee J, Schilbach S, **Goddard A** and Dixit V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J. Biol. Chem.* **274**(15): 10287-10292.
- Gurney AL, Marsters SA, Huang RM, Pitti RM, Mark DT, Baldwin DT, Gray AM, Dowd P, Brush J, Heldens S, Schow P, **Goddard AD**, Wood WI, Baker KP, Godowski PJ and Ashkenazi A. (1999) Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Current Biology* **9**(4): 215-218.
- Ridgway JBB, Ng E, Kern JA, Lee J, Brush J, **Goddard A** and Carter P. (1999) Identification of a human anti-CD55 single-chain Fv by subtractive panning of a phage library using tumor and nontumor cell lines. *Cancer Research* **59**: 2718-2723.
- Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT, Godowski PJ, Wood WI, Gurney AL, Hillan KJ, Cohen RL, **Goddard AD**, Botstein D and Ashkenazi A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**(6712): 699-703.
- Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, **Goddard AD**, Hillan KJ, Gurney AL, Botstein D and Levine AJ. (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc. Natl. Acad. Sci. USA.* **95**(25): 14717-14722.
- Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, **Goddard A**, Wood WI, Gurney AL and Godowski PJ. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* **395**(6699): 284-288.
- Merchant AM, Zhu Z, Yuan JQ, **Goddard A**, Adams CW, Presta LG and Carter P. (1998) An efficient route to human bispecific IgG. *Nature Biotechnology* **16**(7): 677-681.
- Marsters SA, Sheridan JP, Pitti RM, Brush J, **Goddard A** and Ashkenazi A. (1998) Identification of a ligand for the death-domain-containing receptor Apo3. *Current Biology* **8**(9): 525-528.
- Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, **Goddard A**, Rosenthal A, Epstein EH Jr. and de Sauvage FJ. (1998) Activating Smoothed mutations in sporadic basal-cell carcinoma. *Nature.* **391**(6662): 90-92.
- Marsters SA, Sheridan JP, Pitti RM, Huang A, Skubatch M, Baldwin D, Yuan J, Gurney A, **Goddard AD**, Godowski P and Ashkenazi A. (1997) A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Current Biology.* **7**(12): 1003-1006.
- Hynes M, Stone DM, Dowd M, Pitts-Meek S, **Goddard A**, Gurney A and Rosenthal A. (1997) Control of cell pattern in the neural tube by the zinc finger transcription factor *Gli-1*. *Neuron* **19**: 15-26.
- Sheridan JP, Marsters SA, Pitti RM, Gurney A., Skubatch M, Baldwin D, Ramakrishnan L, Gray CL, Baker K, Wood WI, **Goddard AD**, Godowski P, and Ashkenazi A. (1997) Control of TRAIL-Induced Apoptosis by a Family of Signaling and Decoy Receptors. *Science* **277** (5327): 818-821.

Goddard AD, Dowd P, Chernausk S, Geffner M, Gertner J, Hintz R, Hopwood N, Kaplan S, Plotnick L, Rogol A, Rosenfield R, Saenger P, Mauras N, Hershkopf R, Angulo M and Attie, K. (1997) Partial growth hormone insensitivity: The role of growth hormone receptor mutations in idiopathic short stature. *J. Pediatr.* **131**: S51-55.

Klein RD, Sherman D, Ho WH, Stone D, Bennett GL, Moffat B, Vandlen R, Simmons L, Gu Q, Hongo JA, Devaux B, Poulsen K, Armanini M, Nozaki C, Asai N, **Goddard A**, Phillips H, Henderson CE, Takahashi M and Rosenthal A. (1997) A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. *Nature*. **387**(6634): 717-21.

Stone DM, Hynes M, Armanini M, Swanson TA, Gu Q, Johnson RL, Scott MP, Pennica D, **Goddard A**, Phillips H, Noll M, Hooper JE, de Sauvage F and Rosenthal A. (1996) The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**(6605): 129-34.

Marsters SA, Sheridan JP, Donahue CJ, Pitti RM, Gray CL, **Goddard AD**, Bauer KD and Ashkenazi A. (1996) Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF-kappa β . *Current Biology* **6**(12): 1669-76.

Rothe M, Xiong J, Shu HB, Williamson K, **Goddard A** and Goeddel DV. (1996) I-TRAF is a novel TRAF-interacting protein that regulates TRAF-mediated signal transduction. *Proc. Natl. Acad. Sci. USA* **93**: 8241-8246.

Yang M, Luoh SM, **Goddard A**, Reilly D, Henzel W and Bass S. (1996) The bglX gene located at 47.8 min on the Escherichia coli chromosome encodes a periplasmic beta-glucosidase. *Microbiology* **142**: 1659-65.

Goddard AD and Black DM. (1996) Familial Cancer in Molecular Endocrinology of Cancer. Waxman, J. Ed. Cambridge University Press, Cambridge UK, pp.187-215.

Treanor JJS, Goodman L, de Sauvage F, Stone DM, Poulson KT, Beck CD, Gray C, Armanini MP, Pollocks RA, Hefti F, Phillips HS, **Goddard A**, Moore MW, Buj-Bello A, Davis AM, Asai N, Takahashi M, Vandlen R, Henderson CE and Rosenthal A. (1996) Characterization of a receptor for GDNF. *Nature* **382**: 80-83.

Klein RD, Gu Q, **Goddard A** and Rosenthal A. (1996) Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. USA* **93**: 7108-7113.

Winslow JW, Moran P, Valverde J, Shih A, Yuan JQ, Wong SC, Tsai SP, **Goddard A**, Henzel WJ, Hefti F and Caras I. (1995) Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in axon bundle formation. *Neuron* **14**: 973-981.

Bennett BD, Zeigler FC, Gu Q, Fendly B, **Goddard AD**, Gillett N and Matthews W. (1995) Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kinase Htk. *Proc. Natl. Acad. Sci. USA* **92**: 1866-1870.

Huang X, Yuang J, **Goddard A**, Foulis A, James RF, Lernmark A, Pujol-Borrell R, Rabinovitch A, Somoza N and Stewart TA. (1995) Interferon expression in the pancreases of patients with type I diabetes. *Diabetes* **44**: 658-664.

Goddard AD, Yuan JQ, Fairbairn L, Dexter M, Borrow J, Kozak C and Solomon E. (1995) Cloning of the murine homolog of the leukemia-associated PML gene. *Mammalian Genome* **6**: 732-737.

Goddard AD, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMTI and The Growth Hormone Insensitivity Study Group. (1995) Mutations of the growth hormone receptor in children with idiopathic short stature. *N. Engl. J. Med.* **333**: 1093-1098.

Kuo SS, Moran P, Gripp J, Armanini M, Phillips HS, **Goddard A** and Caras IW. (1994) Identification and characterization of Batk, a predominantly brain-specific non-receptor protein tyrosine kinase related to Csk. *J. Neurosci. Res.* **38**: 705-715.

Mark MR, Scadden DT, Wang Z, Gu Q, **Goddard A** and Godowski PJ. (1994) Rse, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**: 10720-10728.

Borrow J, Shipley J, Howe K, Kiely F, **Goddard A**, Sheer D, Srivastava A, Antony AC, Fioretos T, Mitelman F and Solomon E. (1994) Molecular analysis of simple variant translocations in acute promyelocytic leukemia. *Genes Chromosomes Cancer* **9**: 234-243.

Goddard AD and Solomon E. (1993) Genetics of Cancer. *Adv. Hum. Genet.* **21**: 321-376.

Borrow J, **Goddard AD**, Gibbons B, Katz F, Swirsky D, Fioretos T, Dube I, Winfield DA, Kingston J, Hagemeijer A, Rees JKH, Lister AT and Solomon E. (1992) Diagnosis of acute promyelocytic leukemia by RT-PCR: Detection of *PML-RARA* and *RARA-PML* fusion transcripts. *Br. J. Haematol.* **82**: 529-540.

Goddard AD, Borrow J and Solomon E. (1992) A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. *Leukemia* **6 Suppl 3**: 117S-119S.

Zhu X, Dunn JM, **Goddard AD**, Squire JA, Becker A, Phillips RA and Gallie BL. (1992) Mechanisms of loss of heterozygosity in retinoblastoma. *Cytogenet. Cell. Genet.* **59**: 248-252.

Foulkes W, **Goddard A.** and Patel K. (1991) Retinoblastoma linked with Seascale [letter]. *British Med. J.* **302**: 409.

Goddard AD, Borrow J, Freemont PS and Solomon E. (1991) Characterization of a novel zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science* **254**: 1371-1374.

Solomon E, Borrow J and **Goddard AD**. (1991) Chromosomal aberrations in cancer. *Science* **254**: 1153-1160.

Pajunen L, Jones TA, **Goddard A**, Sheer D, Solomon E, Pihlajaniemi T and Kivirikko KI. (1991) Regional assignment of the human gene coding for a multifunctional peptide (P4HB) acting as the β -subunit of prolyl-4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenet. Cell. Genet.* **56**: 165-168.

Borrow J, Black DM, **Goddard AD**, Yagle MK, Frischauf A.-M and Solomon E. (1991) Construction and regional localization of a NotI linking library from human chromosome 17q. *Genomics* **10**: 477-480.

Borrow J, **Goddard AD**, Sheer D and Solomon E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Myers JC, Jones TA, Pohjolainen E-R, Kadri AS, **Goddard AD**, Sheer D, Solomon E and Pihlajaniemi T. (1990) Molecular cloning of 5(IV) collagen and assignment of the gene to the region of the X-chromosome containing the Alport Syndrome locus. *Am. J. Hum. Genet.* **46**: 1024-1033.

Gallie BL, Squire JA, **Goddard A**, Dunn JM, Canton M, Hinton D, Zhu X and Phillips RA. (1990) Mechanisms of oncogenesis in retinoblastoma. *Lab. Invest.* **62**: 394-408.

Goddard AD, Phillips RA, Greger V, Passarge E, Hopping W, Gallie BL and Horsthemke B. (1990) Use of the RB1 cDNA as a diagnostic probe in retinoblastoma families. *Clinical Genetics* **37**: 117-126.

Zhu XP, Dunn JM, Phillips RA, **Goddard AD**, Paton KE, Becker A and Gallie BL. (1989) Germline, but not somatic, mutations of the RB1 gene preferentially involve the paternal allele. *Nature* **340**: 312-314.

Gallie BL, Dunn JM, **Goddard A**, Becker A and Phillips RA. (1988) Identification of mutations in the putative retinoblastoma gene. In Molecular Biology of The Eye: Genes, Vision and Ocular Disease. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 88. J. Piatigorsky, T. Shinohara and P.S. Zelenka, Eds. Alan R. Liss, Inc., New York, 1988, pp. 427-436.

Goddard AD, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA and Gallie BL. (1988) Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. *Mol. Cell. Biol.* **8**: 2082-2088.

Squire J, Dunn J, **Goddard A**, Hoffman T, Musarella M, Willard HF, Becker AJ, Gallie BL and Phillips RA. (1986) Cloning of the esterase D gene: A polymorphic gene probe closely linked to the retinoblastoma locus on chromosome 13. *Proc. Natl. Acad. Sci. USA* **83**: 6573-6577.

Squire J, **Goddard AD**, Canton M, Becker A, Phillips RA and Gallie BL (1986) Tumour induction by the retinoblastoma mutation is independent of N-myc expression. *Nature* **322**: 555-557.

Goddard AD, Heddle JA, Gallie BL and Phillips RA. (1985) Radiation sensitivity of fibroblasts of bilateral retinoblastoma patients as determined by micronucleus induction *in vitro*. *Mutation Research* **152**: 31-38.

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger¹, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. ¹Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR¹ to clinical diagnostics are well known^{2,3}, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases⁴ made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization^{5,6}, gel electrophoresis with or without use of restriction digestion^{7,8}, HPLC⁹, or capillary electrophoresis¹⁰. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing¹¹.

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.¹², developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.¹³, developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to ds-DNA¹⁴⁻¹⁶. As outlined in Figure 1, a prototypic PCR

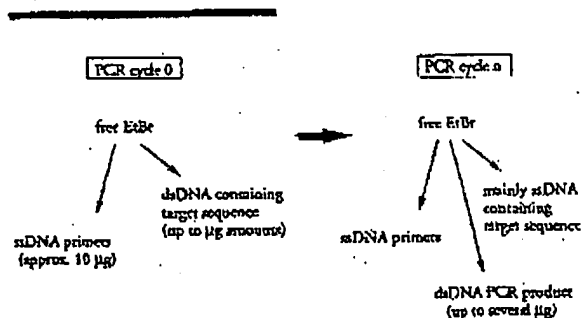


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

otoxin fusion
Sci. USA 89:

drano, T. A.,
immunoprotein
to *Pseudomonas*

ngham, M. C.,
od of cloning
in *es* as single-
1066-1070.
schberg, D. L.,
our effects of
topoxide phos-

r, G., Delcide,
penetra, A. A.
re-targeted by
1184-1189.
et al., Vol. 2, p.

r, and Stevens,
with anti-viral
J. Biochem. J.

A. L., Carnicelli,
d properties of
oxidase activity.

rization of the
abits eukaryotic
528.
urification and
theca americana
chem. Biophys.

1982. Purifica-
of the antiviral
"okweed". Bio-

I. Dodecantrin,
dodecantrin. Bio-

new inhibitor of
tem. 255:6947-

Abbondanza, A.,
ribosome-inacti-
(white bryony).

-synthesis inhib-
Lett. 153:209-

8. Isolation and
inhibitory pro-
them. 52:1223-

v, L., Sperti, S.
also by proteins
clon). Biochem.

nza, A., Cenzori,
Purification and
ith RNA N-gly-
ylation from the
Acta. 993:287-

Guillemet, J. C.,
1988. Trichokl-
of *Trichostema*
8.
itors of animal
Biophys. Acta

roperties of the
oicin inhibiting

of abrin: a toxic
ferent biological
ur. J. Biochem.

Frauz, M. 1980.
i *Vicium altum* L.

i. and Stirpe, F.
if modocia, the

., Brown, A. N.,
s of volkensin, a
14589-14593.
nd properties of
natry 18:2615-

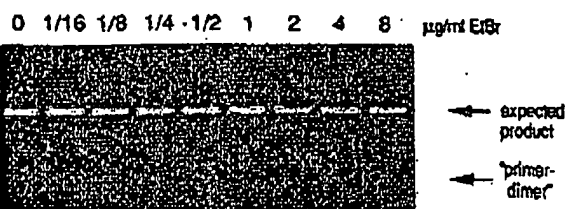
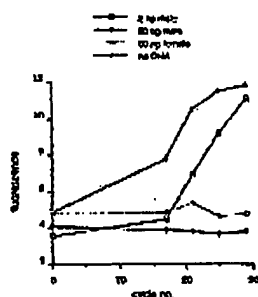


FIGURE 2 Gel electrophoresis of PCR amplification products of the human nuclear gene, HLA DQ α , made in the presence of increasing amounts of EtBr (up to 8 μ g/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

A.



B.

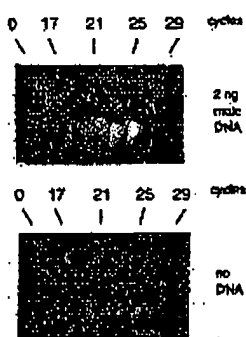


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 μ g/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photography of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ss-DNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA¹⁷ to micrograms per PCR¹⁸. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the affect of EtBr in PCR, amplifications of the human HLA DQ α gene¹⁹ were performed with the dye present at concentrations from 0.06 to 8.0 μ g/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μ g/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μ g/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome²⁰. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β -globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μ g/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β -globin gene²¹. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β -globin allele present^{21,22}.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type β -globin individual (AA); from a heterozygous sickle β -globin individual (AS); and from a homozygous sickle β -globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

emocy.

ess the
HLA
cent at
oncen-
lowing
e 2, gel
ie yield
Br was
ndicat.

Se se-
nent of
ries of
rimers
human
either
DNA.
After 0,
each
is fluo-
plotted
of this
case in
DNA is
umber.
cc-fold
ontain-
crease
her no
DNA
fewer
in fluo-
f these
the ex-
taining
in the

ualized
n a UV
h a red
ns that
VA.

-globin
sch has
etection

Figure
ications
ography
These
for ci-
human
nparted
anal 3'
primer
has am-
c of the
ent^{21,22}

nsists of
the (left
). Three
zygous,
ozygous
zygous
genomic
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells⁶. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

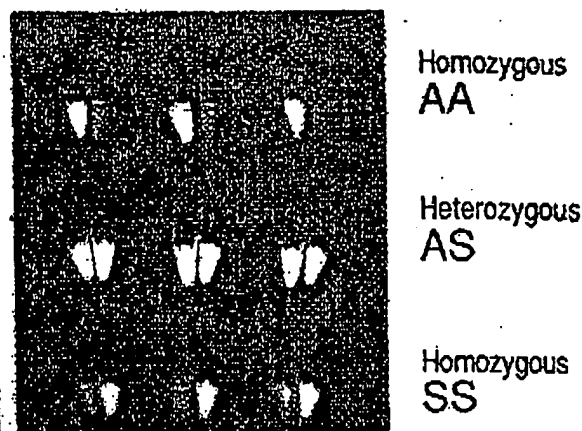


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

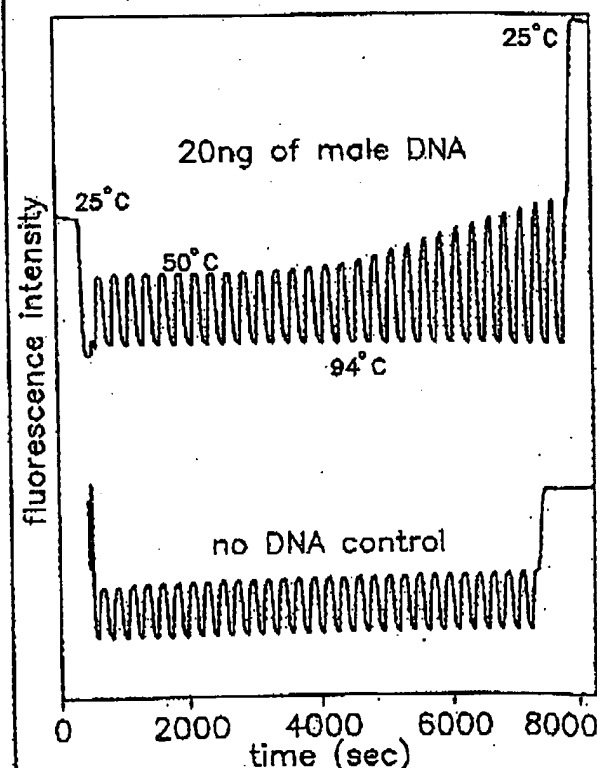


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube³, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins²³. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10⁵ cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer²⁴.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format²⁵. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader²⁶.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptic transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQ α gene amplifications containing EtBr. PCRs were set up in 100 μ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl₂; 2.5 units of Taq DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ α gene specific oligonucleotide primers GH26 and GH27²⁷ and approximately 10³ copies of DQ α PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

Y-chromosome specific PCR. PCRs (100 μ l total reaction volume) containing 0.5 μ g/ml EtBr were prepared as described for HLA-DQ α , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2²⁸ and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β -globin gene PCR. Amplifications of 100 μ l volume using 0.5 μ g/ml of EtBr were prepared as described for HLA-DQ α above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.²¹. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.²¹ to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San-Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griest, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

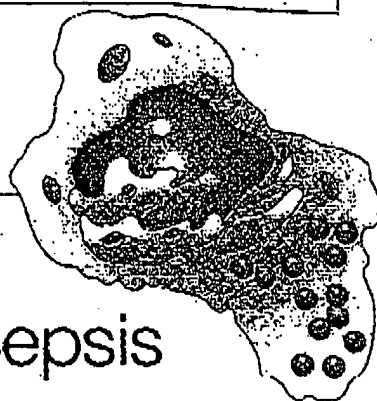
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQ* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Spinsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligonucleotide cleavage detection. *J. Virol.* 61:1690-1694.
- Chhab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:293-294.
- Horn, G. T., Richards, B. and Khugler, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Helger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chhab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Froc, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Sier, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ¹H NMR spectroscopy. *Nuc. Acids Res.* 18:3752-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 334:414-417.
- Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Snijsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. G., Doherty, M. and Giachieri, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Ugazotti, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Snijsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Submitted.*
- Higuchi, R. 1989. Using PCR to engineer DNA. p. 61-70. *In: PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Pionta, E., Williams, J. F. and Wondenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-105, 106-112.
- Tumosa, N. and Kalan, L. 1989. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Meth.* 116:59-63.

IBL

IMMUNO BIOLOGICAL LABORATORIES

sCD-14 ELISA

Trauma, Shock and Sepsis



The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations. During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features: 12x8 determinations (microtiter strips), precoated with a specific monoclonal antibody, 2x1 hour incubation, standard range: 3 - 96 ng/ml detection limit: 1 ng/ml CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND -BIOLOGIE MBH
OSTERSTRASSE 86 · D · 2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98
BIOTECHNOLOGY VOL 10 APRIL 1992

417

GENENTECH, INC.
1 DNA Way
South San Francisco, CA 94080 USA
Phone: (650) 225-1000

FAX: (650) 952-9881

FACSIMILE TRANSMITTAL

Date: 19 July 2004

To: Anna Barry
Heller Ehrman

Re: Higuchi reference

Fax No: 324-6638

From: Patty Tobin, Assistant to Elizabeth M. Barnes, Ph.D.
Genentech, Inc. Legal Department

Number of Pages including this cover sheet: 6

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger¹, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. ¹Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR¹ to clinical diagnostics are well known^{2,3}, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases⁴ made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization^{5,6}, gel electrophoresis with or without use of restriction digestion^{7,8}, HPLC⁹, or capillary electrophoresis¹⁰. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing¹¹.

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.¹², developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.¹³, developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to ds-DNA¹⁴⁻¹⁶. As outlined in Figure 1, a prototypic PCR

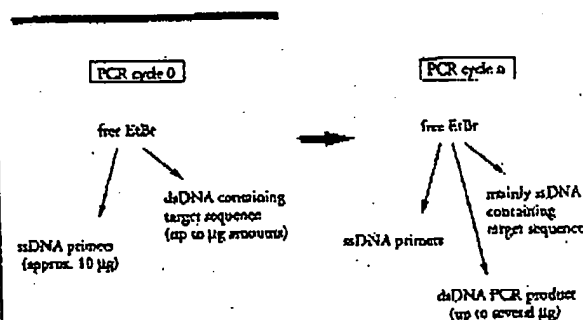


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

otoxin fusion
Sci. USA 85,
dman, T. A.,
immunoproteas-
to *Pseudomonas*
ngam, M. G.,
od of cloning
in *esb* as single
1066-1070.
schberg, D. L.,
our effects of
topoxide phos-
G. Delcide,
penetia, A. A.
re-targeted by
1184-1188.
ed, Vol. 2, p.
and Stevens,
with anti-viral
J. Biochem. J.
A. L. Carnicelli,
d properties of
oxidase activity.
ization of the
abius eukaryotic
528.
urification and
place *arabidopsis*
them. Biophys.
1982. Purifica-
of the antiviral
bikeweed). Bio-
L. Dodecandrin,
dodecandrin. Bio-
new inhibitor of
rent. 255:6947-
Abbondanza, A.,
ribosome-inacti-
(white bryony).
-synthesis inhibi-
Lett. 153:209-
S. Inactivation and
inhibitory pro-
them. 52:1223-
o, L., Sperti, S.
ure by protease
elom). Biochem.
nza, A., Cesini,
Purification and
ith RNA N-gly-
sation from the
Acta. 990:387-
Guillemot, J. C.,
1988. Trichoki-
of *Trichostema*
8.
itors of animal
Biophys. Acta
roperties of the
vacin inhibiting
of abrin: a toxic
ferent biological
ur. J. Biochem.
Franz, H. 1980.
i *Vicia alba* L.
i. and Strpe, F.
of *modestia*, the
i., Brown, A. N.,
s of volkensin, a
14589-14595.
nd properties of
nlutry 18:2615-

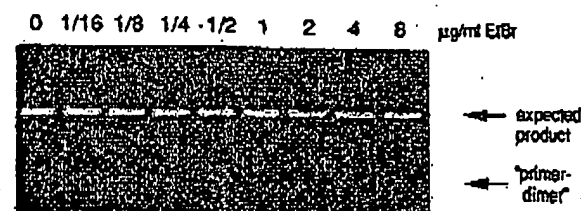


FIGURE 2 Gel electrophoresis of PCR amplification products of the human, nuclear gene, HLA DQ α , made in the presence of increasing amounts of EtBr (up to 8 μ g/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

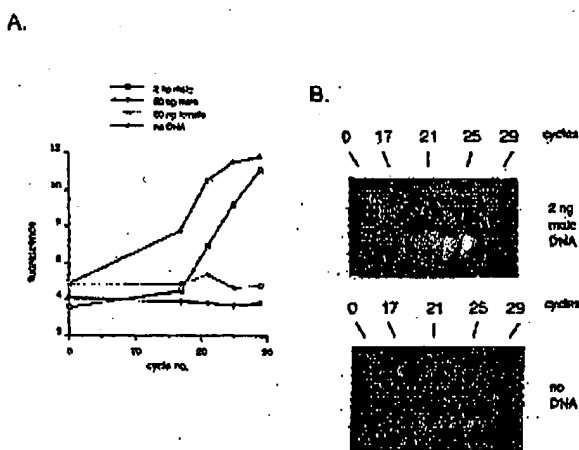


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 μ g/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photograph of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ss-DNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA¹⁷ to micrograms per PCR¹⁸. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the effect of EtBr in PCR, amplifications of the human HLA DQ α gene¹⁹ were performed with the dye present at concentrations from 0.06 to 8.0 μ g/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μ g/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μ g/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome²⁰. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β -globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μ g/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β -globin gene²¹. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β -globin allele present^{21,22}.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type β -globin individual (AA); from a heterozygous sickle β -globin individual (AS); and from a homozygous sickle β -globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

rmocy.

ess the
1 HLA
cent at
oncen-
lowing
e 2, gel
ic yield
Br was
indicat.

Se se-
nent of
ries of
rimers
human
either
DNA.
After 0,
or each
is fluo-
rized
of this
case in
DNA is
umber.
cc-fold
ontain-
crease
her no
DNA
fewer
in fluo-
f these
the ex-
taining
in the

ualized
n a UV
h a red
ms that
NA.
-globin
sch has
etection
Figure
ications
graphy
These
for ci-
human
nparted
ual 3'
primer
has am-
c of the
ent^{21,22}
nsists of
the (left
Three
zygous,
ozygous
ozygous
genomic
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells⁶. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

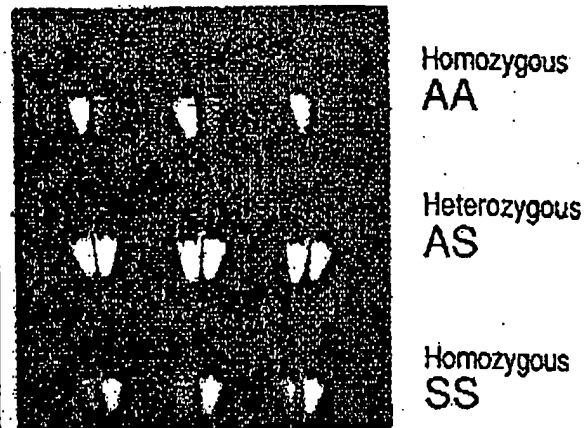


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

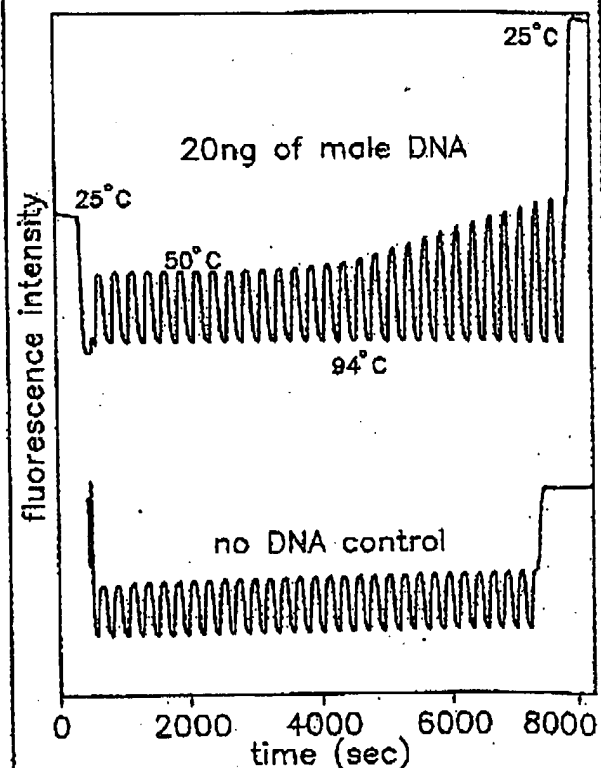


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light in a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube³, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins²³. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^5 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer²⁴.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format²⁵. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader²⁶.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQ α gene amplifications containing EtBr. PCRs were set up in 100 μ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl₂; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ α gene specific oligonucleotide primers (GH26 and GH27)¹⁹ and approximately 10^3 copies of DQ α PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

Y-chromosome specific PCR. PCRs (100 μ l total reaction volume) containing 0.5 μ g/ml EtBr were prepared as described for HLA-DQ α , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2²⁰, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β -globin gene PCR. Amplifications of 100 μ l volume using 0.5 μ g/ml of EtBr were prepared as described for HLA-DQ α above except with different primers and target DNAs. These PCRs contained either primer pair HGP1/H β 14A (wild-type globin specific primers) or HGP2/H β 14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.²¹. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.²¹ to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San-Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griest, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQB* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Sninsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:467-471.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, C. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. *J. Virol.* 61:1690-1694.
- Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:293-294.
- Horn, G. T., Richards, B. and Klinger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Helger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Sacz, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ¹H NMR spectroscopy. *Nuc. Acids Res.* 18:3755-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:414-417.
- Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Sninsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. C., Doherty, M. and Giachier, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Ugazochi, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spasek, D., Gada, L., Levenson, C. and Sninsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Submitted.
- Higuchi, R. 1989. Using PCR to engineer DNA. p. 61-70. In: *PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Pionta, E., Williams, J. F. and Wondenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-105, 106-112.
- Tumosa, N. and Kalish, L. 1989. Fluorescent ELISA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Meth.* 116:59-63.

IBL

IMMUNO BIOLOGICAL LABORATORIES

sCD-14 ELISA

Trauma, Shock and Sepsis

The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations.

During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features: 12x8 determinations (microtiter strips), precoated with a specific monoclonal antibody, 2x1 hour incubation, standard range: 3 - 96 ng/ml detection limit: 1 ng/ml CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND - BIOLOGIE MBH
OSTERSTRASSE 86 · D · 2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98

BIOTECHNOLOGY VOL 10 APRIL 1992

417

Research

Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization

Kenneth J. Livak, Susan J.A. Flood, Jeffrey Marmaro, William Giusti, and Karin Deetz

Perkin-Elmer, Applied Biosystems Division, Foster City, California 94404

The 5' nuclease PCR assay detects the accumulation of specific PCR product by hybridization and cleavage of a double-labeled fluorogenic probe during the amplification reaction. The probe is an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An increase in reporter fluorescence intensity indicates that the probe has hybridized to the target PCR product and has been cleaved by the 5'→3' nucleolytic activity of *Taq* DNA polymerase. In this study, probes with the quencher dye attached to an internal nucleotide were compared with probes with the quencher dye attached to the 3'-end nucleotide. In all cases, the reporter dye was attached to the 5' end. All intact probes showed quenching of the reporter fluorescence. In general, probes with the quencher dye attached to the 3'-end nucleotide exhibited a larger signal in the 5' nuclease PCR assay than the internally labeled probes. It is proposed that the larger signal is caused by increased likelihood of cleavage by *Taq* DNA polymerase when the probe is hybridized to a template strand during PCR. Probes with the quencher dye attached to the 3'-end nucleotide also exhibited an increase in reporter fluorescence intensity when hybridized to a complementary strand. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends can be used as homogeneous hybridization probes.

A homogeneous assay for detecting the accumulation of specific PCR product that uses a double-labeled fluorogenic probe was described by Lee et al.⁽¹⁾ The assay exploits the 5'→3' nucleolytic activity of *Taq* DNA polymerase^(2,3) and is diagramed in Figure 1. The fluorogenic probe consists of an oligonucleotide with a reporter fluorescent dye, such as a fluorescein, attached to the 5' end; and a quencher dye, such as a rhodamine, attached internally. When the fluorescein is excited by irradiation, its fluorescent emission will be quenched if the rhodamine is close enough to be excited through the process of fluorescence energy transfer (FET).^(4,5) During PCR, if the probe is hybridized to a template strand, *Taq* DNA polymerase will cleave the probe because of its inherent 5'→3' nucleolytic activity. If the cleavage occurs between the fluorescein and rhodamine dyes, it causes an increase in fluorescein fluorescence intensity because the fluorescein is no longer quenched. The increase in fluorescein fluorescence intensity indicates that the probe-specific PCR product has been generated. Thus, FET between a reporter dye and a quencher dye is critical to the performance of the probe in the 5' nuclease PCR assay.

Quenching is completely dependent on the physical proximity of the two dyes.⁽⁶⁾ Because of this, it has been assumed that the quencher dye must be attached near the 5' end. Surprisingly, we have found that attaching a rhodamine dye at the 3' end of a probe still provides adequate quenching for the probe to perform in the 5' nuclease

PCR assay. Furthermore, cleavage of this type of probe is not required to achieve some reduction in quenching. Oligonucleotides with a reporter dye on the 5' end and a quencher dye on the 3' end exhibit a much higher reporter fluorescence when double-stranded as compared with single-stranded. This should make it possible to use this type of double-labeled probe for homogeneous detection of nucleic acid hybridization.

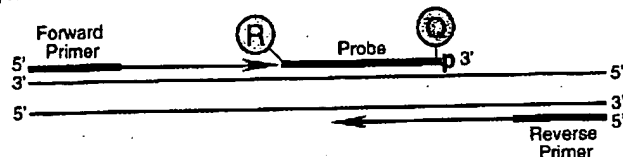
MATERIALS AND METHODS

Oligonucleotides

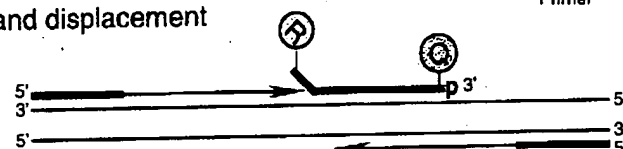
Table 1 shows the nucleotide sequence of the oligonucleotides used in this study. Linker arm nucleotide (LAN) phosphoramidite was obtained from Glen Research. The standard DNA phosphoramidites, 6-carboxyfluorescein (6-FAM) phosphoramidite, 6-carboxytetramethylrhodamine succinimidyl ester (TAMRA NHS ester), and Phosphalink for attaching a 3'-blocking phosphate, were obtained from Perkin-Elmer, Applied Biosystems Division. Oligonucleotide synthesis was performed using an ABI model 394 DNA synthesizer (Applied Biosystems). Primer and complement oligonucleotides were purified using Oligo Purification Cartridges (Applied Biosystems). Double-labeled probes were synthesized with 6-FAM-labeled phosphoramidite at the 5' end, LAN replacing one of the T's in the sequence, and Phosphalink at the 3' end. Following deprotection and ethanol precipitation, TAMRA NHS ester was coupled to the LAN-containing oligonucleotide in 250

Research

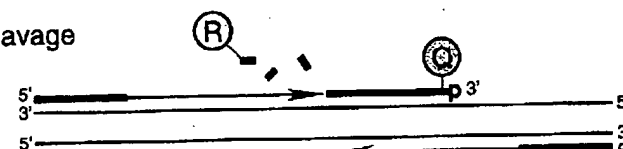
Polymerization



Strand displacement



Cleavage



Polymerization completed

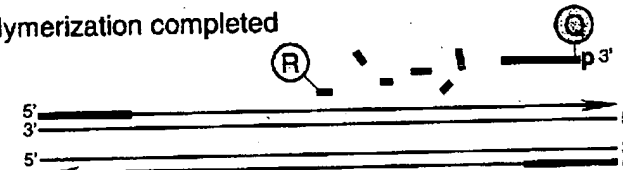


FIGURE 1 Diagram of 5' nuclease assay. Stepwise representation of the 5' → 3' nucleolytic activity of Taq DNA polymerase acting on a fluorogenic probe during one extension phase of PCR.

mm Na-bicarbonate buffer (pH 9.0) at room temperature. Unreacted dye was removed by passage over a PD-10 Sephadex column. Finally, the double-labeled probe was purified by preparative high-performance liquid chromatography (HPLC) using an Aquapore C₈ 220×4.6-mm column with 7-μm particle size. The column was developed with a 24-min linear gradient of 8–20% acetonitrile in 0.1 M TEAA (triethylamine acetate). Probes are named by designating the sequence from Table 1 and the position of the LAN-TAMRA moiety. For example, probe A1-7 has sequence A1 with LAN-TAMRA at nucleotide position 7 from the 5' end.

PCR Systems

All PCR amplifications were performed in the Perkin-Elmer GeneAmp PCR System 9600 using 50-μl reactions that contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 0.5 unit of AmpErase uracil N-glycosylase (Perkin-Elmer), and 1.25 unit of AmpliTaq DNA polymerase (Perkin-Elmer). A 295-bp segment from exon 3 of the human β-actin

gene (nucleotides 2141–2435 in the sequence of Nakajima-Iijima et al.⁽⁷⁾) was amplified using primers AFP and ARP (Table 1), which are modified slightly from those of du Breuil et al.⁽⁸⁾ Actin amplification reactions contained 4 mM MgCl₂, 20 ng of human genomic DNA, 50 nM A1 or A3 probe, and 300 nM each

primer. The thermal regimen was 50°C (2 min), 95°C (10 min), 40 cycles of 95°C (20 sec), 60°C (1 min), and hold at 72°C. A 515-bp segment was amplified from a plasmid that consists of a segment of λ DNA (nucleotides 32,220–32,747) inserted in the *Sma*I site of vector pUC119. These reactions contained 3.5 mM MgCl₂, 1 ng of plasmid DNA, 50 nM P2 or P5 probe, 200 nM primer F119, and 200 nM primer R119. The thermal regimen was 50°C (2 min), 95°C (10 min), 25 cycles of 95°C (20 sec), 57°C (1 min), and hold at 72°C.

Fluorescence Detection

For each amplification reaction, a 40-μl aliquot of a sample was transferred to an individual well of a white, 96-well microtiter plate (Perkin-Elmer). Fluorescence was measured on the Perkin-Elmer TaqMan LS-50B System, which consists of a luminescence spectrometer with plate reader assembly, a 485-nm excitation filter, and a 515-nm emission filter. Excitation was at 488 nm using a 5-nm slit width. Emission was measured at 518 nm for 6-FAM (the reporter or R value) and 582 nm for TAMRA (the quencher or Q value) using a 10-nm slit width. To determine the increase in reporter emission that is caused by cleavage of the probe during PCR, three normalizations are applied to the raw emission data. First, emission intensity of a buffer blank is subtracted for each wavelength. Second, emission intensity of the reporter is

TABLE 1 Sequences of Oligonucleotides

Name	Type	Sequence
F119	primer	ACCCACAGGAAGTATCACCCTC
R119	primer	ATGTCGCGTTCCGGCTGACGTTCTGC
P2	probe	TCGCATTACTGATCGTTCGCCAACCAGTp
P2C	complement	GTACTGGTTGGCAACGATCAGTAATGCCGATG
P5	probe	CGGATTTGCTGGTATCTATGACAAGGATp
P5C	complement	TTCATCCTTGTCTATAGATACCAAGCAATCCG
AFP	primer	TCACCCACACTGTGCCCATCTACGA
ARP	primer	CAGCGGAACCGCTCATTGCCAATGG
A1	probe	ATGCCCTCCCCATGCCATCCTTCGCTp
A1C	complement	AGACGCAGGATGGCATGGGGGAGGGCATA
A3	probe	CGCCCCTGGAAGTTCGAGCAAGAGATp
A3C	complement	CCATCTCTTGCTCGAAGTCCAGGGCGAC

For each oligonucleotide used in this study, the nucleic acid sequence is given, written in the 5' → 3' direction. There are three types of oligonucleotides: PCR primer, fluorogenic probe used in the 5' nuclease assay, and complement used to hybridize to the corresponding probe. For the probes, the underlined base indicates a position where LAN with TAMRA attached was substituted for a T. (p) The presence of a 3' phosphate on each probe.

A1-2 RAQGCCCTCCCCATGCCATCCTGCGTp
 A1-7 RATGCCCQCCCCATGCCATCCTGCGTp
 A1-14 RATGCCCTCCCCCAQGCCATCCTGCGTp
 A1-19 RATGCCCTCCCCATGCCAQCTGCGTp
 A1-22 RATGCCCTCCCCATGCCATCCQCGTp
 A1-26 RATGCCCTCCCCATGCCATCCTGCGQp

Probe	518 nm		582 nm		RQ ⁻	RQ ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A1-2	25.5 ± 2.1	32.7 ± 1.9	38.2 ± 3.0	38.2 ± 2.0	0.67 ± 0.01	0.86 ± 0.06	0.19 ± 0.06
A1-7	53.5 ± 4.3	395.1 ± 21.4	108.5 ± 6.3	110.3 ± 5.3	0.49 ± 0.03	3.58 ± 0.17	3.09 ± 0.18
A1-14	127.0 ± 4.9	403.5 ± 19.1	109.7 ± 5.3	93.1 ± 6.3	1.16 ± 0.02	4.34 ± 0.15	3.18 ± 0.15
A1-19	187.5 ± 17.9	422.7 ± 7.7	70.3 ± 7.4	73.0 ± 2.8	2.87 ± 0.05	5.80 ± 0.15	3.13 ± 0.16
A1-22	224.6 ± 9.4	482.2 ± 43.6	100.0 ± 4.0	96.2 ± 9.6	2.25 ± 0.03	5.02 ± 0.11	2.77 ± 0.12
A1-26	160.2 ± 8.9	454.1 ± 18.4	93.1 ± 5.4	90.7 ± 3.2	1.72 ± 0.02	5.01 ± 0.08	3.29 ± 0.08

FIGURE 2 Results of 5' nuclease assay comparing β-actin probes with TAMRA at different nucleotide positions. As described in Materials and Methods, PCR amplifications containing the indicated probes were performed, and the fluorescence emission was measured at 518 and 582 nm. Reported values are the average ± 1 s.d. for six reactions run without added template (no temp.) and six reactions run with template (+ temp.). The RQ ratio was calculated for each individual reaction and averaged to give the reported RQ⁻ and RQ⁺ values.

divided by the emission intensity of the quencher to give an RQ ratio for each reaction tube. This normalizes for well-to-well variations in probe concentration and fluorescence measurement. Finally, ΔRQ is calculated by subtracting the RQ value of the no-template control (RQ⁻) from the RQ value for the complete reaction including template (RQ⁺).

RESULTS

A series of probes with increasing distances between the fluorescein reporter and rhodamine quencher were tested to investigate the minimum and maximum spacing that would give an acceptable performance in the 5' nuclease PCR assay. These probes hybridize to a target

sequence in the human β-actin gene. Figure 2 shows the results of an experiment in which these probes were included in PCR that amplified a segment of the β-actin gene containing the target sequence. Performance in the 5' nuclease PCR assay is monitored by the magnitude of ΔRQ, which is a measure of the increase in reporter fluorescence caused by PCR amplification of the probe target. Probe A1-2 has a ΔRQ value that is close to zero, indicating that the probe was not cleaved appreciably during the amplification reaction. This suggests that with the quencher dye on the second nucleotide from the 5' end, there is insufficient room for *Taq* polymerase to cleave efficiently between the reporter and quencher. The other five probes exhibited comparable ΔRQ values that are

clearly different from zero. Thus, all five probes are being cleaved during PCR amplification resulting in a similar increase in reporter fluorescence. It should be noted that complete digestion of a probe produces a much larger increase in reporter fluorescence than that observed in Figure 2 (data not shown). Thus, even in reactions where amplification occurs, the majority of probe molecules remain uncleaved. It is mainly for this reason that the fluorescence intensity of the quencher dye TAMRA changes little with amplification of the target. This is what allows us to use the 582-nm fluorescence reading as a normalization factor.

The magnitude of RQ⁻ depends mainly on the quenching efficiency inherent in the specific structure of the probe and the purity of the oligonucleotide. Thus, the larger RQ⁻ values indicate that probes A1-14, A1-19, A1-22, and A1-26 probably have reduced quenching as compared with A1-7. Still, the degree of quenching is sufficient to detect a highly significant increase in reporter fluorescence when each of these probes is cleaved during PCR.

To further investigate the ability of TAMRA on the 3' end to quench 6-FAM on the 5' end, three additional pairs of probes were tested in the 5' nuclease PCR assay. For each pair, one probe has TAMRA attached to an internal nucleotide and the other has TAMRA attached to the 3' end nucleotide. The results are shown in Table 2. For all three sets, the probe with the 3' quencher exhibits a ΔRQ value that is considerably higher than for the probe with the internal quencher. The RQ⁻ values suggest that differences in quenching are not as great as those observed with some of the A1 probes. These results demonstrate that a quencher dye on the 3' end of an oligonucleotide can quench efficiently the

TABLE 2 Results of 5' Nuclease Assay Comparing Probes with TAMRA Attached to an Internal or 3'-terminal Nucleotide

Probe	518 nm		582 nm		RQ ⁻	RQ ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A3-6	54.6 ± 3.2	84.8 ± 3.7	116.2 ± 6.4	115.6 ± 2.5	0.47 ± 0.02	0.73 ± 0.03	0.26 ± 0.04
A3-24	72.1 ± 2.9	236.5 ± 11.1	84.2 ± 4.0	90.2 ± 3.8	0.86 ± 0.02	2.62 ± 0.05	1.76 ± 0.05
P2-7	82.8 ± 4.4	384.0 ± 34.1	105.1 ± 6.4	120.4 ± 10.2	0.79 ± 0.02	3.19 ± 0.16	2.40 ± 0.16
P2-27	113.4 ± 6.6	555.4 ± 14.1	140.7 ± 8.5	118.7 ± 4.8	0.81 ± 0.01	4.68 ± 0.10	3.88 ± 0.10
PS-10	77.5 ± 6.5	244.4 ± 15.9	86.7 ± 4.3	95.8 ± 6.7	0.89 ± 0.05	2.55 ± 0.06	1.66 ± 0.08
PS-28	64.0 ± 5.2	333.6 ± 12.1	100.6 ± 6.1	94.7 ± 6.3	0.63 ± 0.02	3.53 ± 0.12	2.89 ± 0.13

Reactions containing the indicated probes and calculations were performed as described in Material and Methods and in the legend to Fig. 2.

[REDACTED]

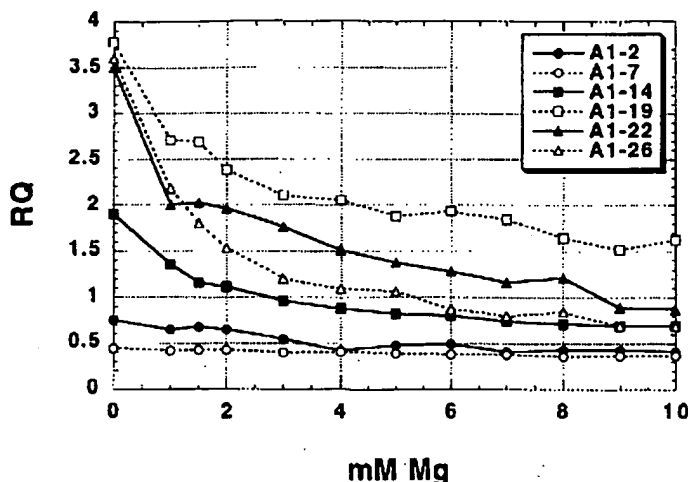


FIGURE 3 Effect of Mg^{2+} concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nM probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of $MgCl_2$. The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs. $MgCl_2$ concentration (mM Mg). The key (upper right) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe T_m , presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.⁽¹⁾

The rise in RQ^- values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3' end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3' end (A1-26). This is understandable, as the conformation of the 3' end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3' end is freer to adopt conformations close to the 5' reporter dye than is an internally placed quencher. For the other three sets of

probes, the interpretation of RQ^- values is less clear-cut. The A3 probes show the same trend as A1, with the 3' TAMRA probe having a larger RQ^- than the internal TAMRA probe. For the P2 pair, both probes have about the same RQ^- value. For the P5 probes, the RQ^- for the 3' probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the RQ^- value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on RQ^- .

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes, ΔRQ is much greater for the 3' TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3' end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as ΔRQ does not increase when the quencher is placed closer to the 3' end. This illus-

trates the importance of being able to use probes with a quencher on the 3' end in the 5' nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A3-6 presumably means the probe is being cleaved 3' to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5' nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3' end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base-pairing and reduce the T_m of a probe. In fact, a 2°C–3°C reduction in T_m has been observed for two probes with internally attached TAMRAs.⁽⁹⁾ This disruptive effect would be minimized by placing the quencher at the 3' end. Thus, probes with 3' quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3' quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.⁽¹¹⁾ demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the $\Delta F508$ mutant. Their probes had TAMRA attached to the seventh nucleotide from

Research

the 5' end and were designed so that any mismatches were between the reporter and quencher. Increasing the distance between reporter and quencher would lessen the disruptive effect of mismatches and allow cleavage of the probe on the incorrect target. Thus, probes with a quencher attached to an internal nucleotide may still be useful for allelic discrimination.

In this study loss of quenching upon hybridization was used to show that quenching by a 3' TAMRA is dependent on the flexibility of a single-stranded oligonucleotide. The increase in reporter fluorescence intensity, though, could also be used to determine whether hybridization has occurred or not. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends should also be useful as hybridization probes. The ability to detect hybridization in real time means that these probes could be used to measure hybridization kinetics. Also, this type of probe could be used to develop homogeneous hybridization assays for diagnostics or other applications. Bagwell et al.⁽¹⁰⁾ describe just this type of homogeneous assay where hybridization of a probe causes an increase in fluorescence caused by a loss of quenching. However, they utilized a complex probe design that requires adding nucleotides to both ends of the probe sequence to form two imperfect hairpins. The results presented here demonstrate that the simple addition of a reporter dye to one end of an oligonucleotide and a quencher dye to the other end generates a fluorogenic probe that can detect hybridization or PCR amplification.

ACKNOWLEDGMENTS

We acknowledge Lincoln McBride of Perkin-Elmer for his support and encouragement on this project and Mitch Winnik of the University of Toronto for helpful discussions on time-resolved fluorescence.

REFERENCES

1. Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* **21**: 3761-3766.
2. Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction prod-

uct by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* **88**: 7276-7280.

3. Lyamichev, V., M.A.D. Brow, and J.E. Dahlberg. 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260**: 778-783.
4. Förster, V.Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* **2**: 55-75.
5. Lakowicz, J.R. 1983. Energy transfer. In *Principles of fluorescent spectroscopy*, pp. 303-339. Plenum Press, New York, NY.
6. Stryer, L. and R.P. Haugland. 1967. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci.* **58**: 719-726.
7. Nakajima-Iijima, S., H. Hamada, P. Reddy, and T. Kakunaga. 1985. Molecular structure of the human cytoplasmic beta-actin gene: Inter-species homology of sequences in the introns. *Proc. Natl. Acad. Sci.* **82**: 6133-6137.
8. du Breuil, R.M., J.M. Patel, and B.V. Mendelow. 1993. Quantitation of β -actin-specific mRNA transcripts using xeno-competitive PCR. *PCR Methods Applic.* **3**: 57-59.
9. Livak, K.J. (unpubl.).
10. Bagwell, C.B., M.E. Munson, R.L. Christensen, and E.J. Lovett. 1994. A new homogeneous assay system for specific nucleic acid sequences: Poly-dA and poly-A detection. *Nucleic Acids Res.* **22**: 2424-2425.

Received December 20, 1994; accepted in revised form March 6, 1995.

THIS MATERIAL MAY BE PROTECTED
BY COPYRIGHT LAW (17 U.S. CODE)

GENOME METHODS

Real Time Quantitative PCR

Christian A. Heid,¹ Junko Stevens,² Kenneth J. Livak,² and
P. Mickey Williams^{1,3}

¹BioAnalytical Technology Department, Genentech, Inc., South San Francisco, California 94080;

²Applied BioSystems Division of Perkin Elmer Corp., Foster City, California 94404

We have developed a novel "real time" quantitative PCR method. The method measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan Probe). This method provides very accurate and reproducible quantitation of gene copies. Unlike other quantitative PCR methods, real-time PCR does not require post-PCR sample handling, preventing potential PCR product carry-over contamination and resulting in much faster and higher throughput assays. The real-time PCR method has a very large dynamic range of starting target molecule determination (at least five orders of magnitude). Real-time quantitative PCR is extremely accurate and less labor-intensive than current quantitative PCR methods.

Quantitative nucleic acid sequence analysis has had an important role in many fields of biological research. Measurement of gene expression (RNA) has been used extensively in monitoring biological responses to various stimuli (Tan et al. 1994; Huang et al. 1995a,b; Prud'homme et al. 1995). Quantitative gene analysis (DNA) has been used to determine the genome quantity of a particular gene, as in the case of the human *HER2* gene, which is amplified in ~30% of breast tumors (Slamon et al. 1987). Gene and genome quantitation (DNA and RNA) also have been used for analysis of human immunodeficiency virus (HIV) burden demonstrating changes in the levels of virus throughout the different phases of the disease (Connor et al. 1993; Platak et al. 1993b; Furtado et al. 1995).

Many methods have been described for the quantitative analysis of nucleic acid sequences (both for RNA and DNA; Southern 1975; Sharp et al. 1980; Thomas 1980). Recently, PCR has proven to be a powerful tool for quantitative nucleic acid analysis. PCR and reverse transcriptase (RT)-PCR have permitted the analysis of minimal starting quantities of nucleic acid (as little as one cell equivalent). This has made possible many experiments that could not have been performed with traditional methods. Although PCR has provided a powerful tool, it is imperative

that it be used properly for quantitation (Ragymaekers 1995). Many early reports of quantitative PCR and RT-PCR described quantitation of the PCR product but did not measure the initial target sequence quantity. It is essential to design proper controls for the quantitation of the initial target sequences (Perre 1992; Clementi et al. 1993).

Researchers have developed several methods of quantitative PCR and RT-PCR. One approach measures PCR product quantity in the log phase of the reaction before the plateau (Kellogg et al. 1990; Pang et al. 1990). This method requires that each sample has equal input amounts of nucleic acid and that each sample under analysis amplifies with identical efficiency up to the point of quantitative analysis. A gene sequence (contained in all samples at relatively constant quantities, such as β -actin) can be used for sample amplification efficiency normalization. Using conventional methods of PCR detection and quantitation (gel electrophoresis or plate capture hybridization), it is extremely laborious to assure that all samples are analyzed during the log phase of the reaction (for both the target gene and the normalization gene). Another method, quantitative competitive (QC)-PCR, has been developed and is used widely for PCR quantitation. QC-PCR relies on the inclusion of an internal control competitor in each reaction (Becker-Andre 1991; Platak et al. 1993a,b). The efficiency of each reaction is normalized to the internal competitor. A known amount of internal competitor can be

³Corresponding author.

REAL TIME QUANTITATIVE PCR

RESULTS

PCR Product Detection in Real Time

added to each sample. To obtain relative quantitation, the unknown target PCR product is compared with the known competitor PCR product. Success of a quantitative competitive PCR assay relies on developing an internal control that amplifies with the same efficiency as the target molecule. The design of the competitor and the validation of amplification efficiencies require a dedicated effort. However, because QC-PCR does not require that PCR products be analyzed during the log phase of the amplification, it is the easier of the two methods to use.

Several detection systems are used for quantitative PCR and RT-PCR analysis: (1) agarose gels, (2) fluorescent labelling of PCR products and detection with laser-induced fluorescence using capillary electrophoresis (Fusco et al. 1995; Williams et al. 1996) or acrylamide gels, and (3) plate capture and sandwich probe hybridization (Mulder et al. 1994). Although these methods proved successful, each method requires post-PCR manipulations that add time to the analysis and may lead to laboratory contamination. The sample throughput of these methods is limited (with the exception of the plate capture approach), and, therefore, these methods are not well suited for uses demanding high sample throughput (i.e., screening of large numbers of biomolecules or analyzing samples for diagnostics or clinical trials).

Here we report the development of a novel assay for quantitative DNA analysis. The assay is based on the use of the 5' nuclease assay first described by Holland et al. (1991). The method uses the 5' nuclease activity of *Taq* polymerase to cleave a nonextendible hybridization probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee et al. 1993; Bassler et al. 1995; Ljvak et al. 1995a,b). One fluorescent dye serves as a reporter [FAM (i.e., 6-carboxyfluorescein)] and its emission spectra is quenched by the second fluorescent dye, TAMRA (i.e., 6-carboxy-tetramethylrhodamine). The nuclease degradation of the hybridization probe releases the quenching of the FAM fluorescent emission, resulting in an increase in peak fluorescent emission at 518 nm. The use of a sequence detector (ABI Prism) allows measurement of fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Therefore, the reactions are monitored in real time. The output data is described and quantitative analysis of input target DNA sequences is discussed below.

The goal was to develop a high-throughput, sensitive, and accurate gene quantitation assay for use in monitoring lipid mediated therapeutic gene delivery. A plasmid encoding human factor VIII gene sequence, pF8TM (see Methods), was used as a model therapeutic gene. The assay uses fluorescent Taqman methodology and an instrument capable of measuring fluorescence in real time (ABI Prism 7700 Sequence Detector). The Taqman reaction requires a hybridization probe labeled with two different fluorescent dyes. One dye is a reporter dye (FAM), the other is a quenching dye (TAMRA). When the probe is intact, fluorescent energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye (TAMRA). During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5'-3' nucleolytic activity of the DNA polymerase. On cleavage of the probe, the reporter dye emission is no longer transferred efficiently to the quenching dye, resulting in an increase of the reporter dye fluorescent emission spectra. PCR primers and probes were designed for the human factor VIII sequence and human β -actin gene (as described in Methods). Optimization reactions were performed to choose the appropriate probe and magnesium concentrations yielding the highest intensity of reporter fluorescent signal without sacrificing specificity. The instrument uses a charge-coupled device (i.e., CCD camera) for measuring the fluorescent emission spectra from 500 to 650 nm. Each PCR tube was monitored sequentially for 25 msec with continuous monitoring throughout the amplification. Each tube was re-examined every 8.5 sec. Computer software was designed to examine the fluorescent intensity of both the reporter dye (FAM) and the quenching dye (TAMRA). The fluorescent intensity of the quenching dye, TAMRA, changes very little over the course of the PCR amplification (data not shown). Therefore, the intensity of TAMRA dye emission serves as an internal standard with which to normalize the reporter dye (FAM) emission variations. The software calculates a value termed ΔRn (or $\Delta R(2)$) using the following equation: $\Delta Rn = (Rn^+) - (Rn^-)$, where Rn^+ = emission intensity of reporter/emission intensity of quencher at any given time in a reaction tube, and Rn^- = emission intensity of re-

HLID ET AL.

porter/emission intensity of quencher measured prior to PCR amplification in that same reaction tube. For the purpose of quantitation, the last three data points (ΔRn s) collected during the extension step for each PCR cycle were analyzed. The nucleolytic degradation of the hybridization probe occurs during the extension phase of PCR, and, therefore, reporter fluorescent emission increases during this time. The three data points were averaged for each PCR cycle and the mean value for each was plotted in an "amplification plot" shown in Figure 1A. The ΔRn mean value is plotted on the y-axis, and time, represented by cycle number, is plotted on the x-axis. During the early cycles of the PCR amplification, the ΔRn

value remains at base line. When sufficient hybridization probe has been cleaved by the *Taq* polymerase nuclease activity, the intensity of reporter fluorescent emission increases. Most PCR amplifications reach a plateau phase of reporter fluorescent emission if the reaction is carried out to high cycle numbers. The amplification plot is examined early in the reaction, at a point that represents the log phase of product accumulation. This is done by assigning an arbitrary threshold that is based on the variability of the base-line data. In Figure 1A, the threshold was set at 10 standard deviations above the mean of base line emission calculated from cycles 1 to 15. Once the threshold is chosen, the point at which

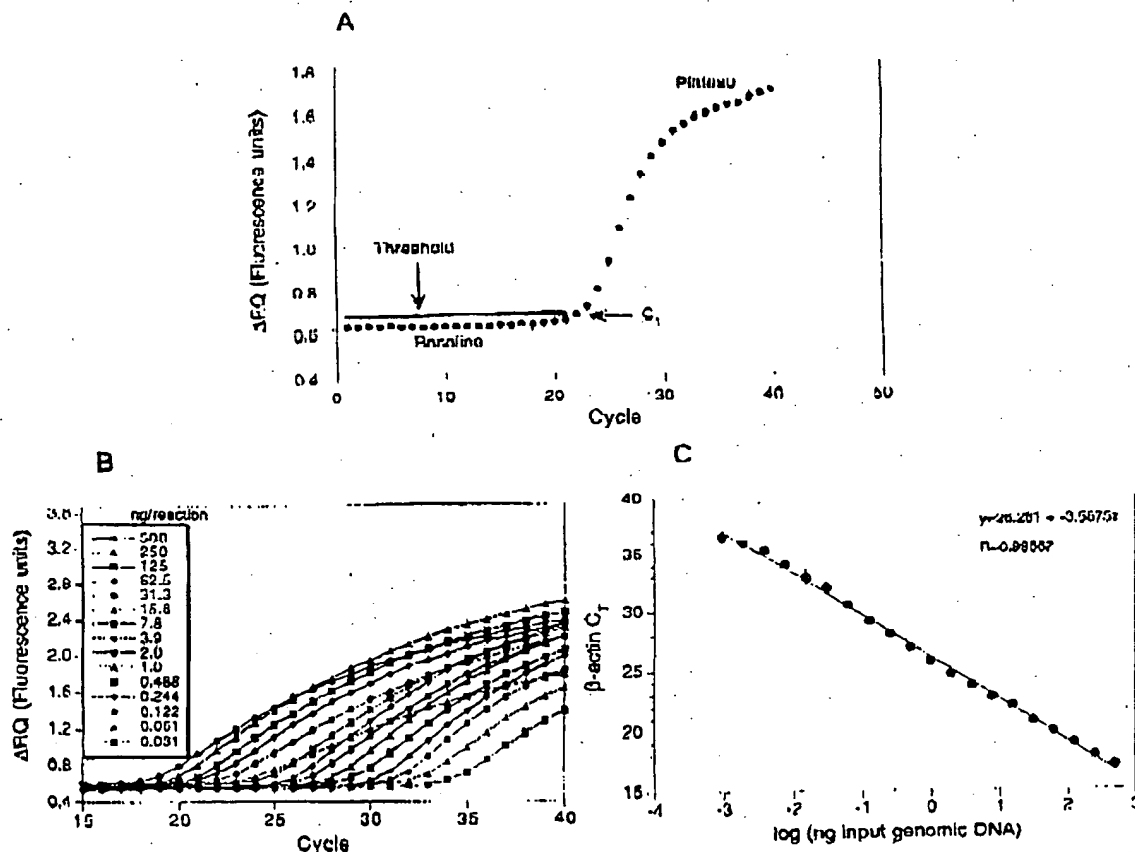


Figure 1 PCR product detection in real time. (A) The Model 7700 software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot. C_T values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line). (B) Overlay of amplification plots of serially (1:2) diluted human genomic DNA samples amplified with β -actin primers. (C) Input DNA concentration of the samples plotted versus C_T . All

REAL TIME QUANTITATIVE PCR

the amplification plot crosses the threshold is defined as C_T . C_T is reported as the cycle number at this point. As will be demonstrated, the C_T value is predictive of the quantity of input target.

C_T Values Provide a Quantitative Measurement of Input Target Sequences

Figure 1B shows amplification plots of 15 different PCR amplifications overlaid. The amplifications were performed on a 1:2 serial dilution of human genomic DNA. The amplified target was human β actin. The amplification plots shift to the right (to higher threshold cycles) as the input target quantity is reduced. This is expected because reactions with fewer starting copies of the target molecule require greater amplification to degrade enough probe to attain the threshold fluorescence. An arbitrary threshold of 10 standard deviations above the base line was used to determine the C_T values. Figure 1C represents the C_T values plotted versus the sample dilution value. Each dilution was amplified in triplicate PCR amplifications and plotted as mean values with error bars representing one standard deviation. The C_T values decrease linearly with increasing target quantity. Thus, C_T values can be used as a quantitative measurement of the input target number. It should be noted that the amplification plot for the 15.6-ng sample shown in Figure 1B does not reflect the same fluorescent rate of increase exhibited by most of the other samples. The 15.6-ng sample also achieves endpoint plateau at a lower fluorescent value than would be expected based on the input DNA. This phenomenon has been observed occasionally with other samples (data not shown) and may be attributable to late cycle inhibition; this hypothesis is still under investigation. It is important to note that the flattened slope and early plateau do not impact significantly the calculated C_T value as demonstrated by the fit on the line shown in Figure 1C. All triplicate amplifications resulted in very similar C_T values—the standard deviation did not exceed 0.5 for any dilution. This experiment contains a >100,000-fold range of input target molecules. Using C_T values for quantitation permits a much larger assay range than directly using total fluorescent emission intensity for quantitation. The linear range of fluorescent intensity measurement of the ABI Prism 7700 Se-

ments over a very large range of relative starting target quantities.

Sample Preparation Validation

Several parameters influence the efficiency of PCR amplification: magnesium and salt concentrations, reaction conditions (i.e., time and temperature), PCR target size and composition, primer sequences, and sample purity. All of the above factors are common to a single PCR assay, except sample to sample purity. In an effort to validate the method of sample preparation for the factor VIII assay, PCR amplification reproducibility and efficiency of 10 replicate sample preparations were examined. After genomic DNA was prepared from the 10 replicate samples, the DNA was quantitated by ultraviolet spectroscopy. Amplifications were performed analyzing β -actin gene content in 100 and 25 ng of total genomic DNA. Each PCR amplification was performed in triplicate. Comparison of C_T values for each triplicate sample show minimal variation based on standard deviation and coefficient of variance (Table 1). Therefore, each of the triplicate PCR amplifications was highly reproducible, demonstrating that real time PCR using this instrumentation introduces minimal variation into the quantitative PCR analysis. Comparison of the mean C_T values of the 10 replicate sample preparations also showed minimal variability, indicating that each sample preparation yielded similar results for β -actin gene quantity. The highest C_T difference between any of the samples was 0.85 and 0.71 for the 100 and 25 ng samples, respectively. Additionally, the amplification of each sample exhibited an equivalent rate of fluorescent emission intensity change per amount of DNA target analyzed as indicated by similar slopes derived from the sample dilutions (Fig. 2). Any sample containing an excess of a PCR inhibitor would exhibit a greater measured β -actin C_T value for a given quantity of DNA. In addition, the inhibitor would be diluted along with the sample in the dilution analysis (Fig. 2), altering the expected C_T value change. Each sample amplification yielded a similar result in the analysis, demonstrating that this method of sample preparation is highly reproducible with regard to sample purity.

Quantitative Analysis of a Plasmid After

HEIDI ET AL

Table 1. Reproducibility of Sample Preparation Method

Sample no.	100 ng				25 ng			
	C _T	mean	standard deviation	CV	C _T	mean	standard deviation	CV
1	18.24 18.23 18.33	18.27	0.06	0.32	20.48 20.55 20.5	20.51	0.03	0.17
2	18.33 18.35 18.44				20.61 20.59 20.41			
3	18.3 18.3 18.42	18.37	0.06	0.32	20.54 20.6 20.49	20.54	0.11	0.54
4	18.15 18.23 18.32				20.48 20.44 20.38			
5	18.4 18.38 18.46	18.34	0.07	0.36	20.68 20.87 20.63	20.54	0.06	0.28
6	18.54 18.67 19				21.09 21.04 21.01			
7	18.28 18.36 18.52	18.71	0.24	1.26	20.67 20.73 20.65	21.06	0.03	0.15
8	18.45 18.7 18.73				20.98 20.84 20.75			
9	18.18 18.34 18.26	18.63	0.16	0.83	20.46 20.54 20.48	20.86	0.12	0.57
10	18.42 18.57 18.66				20.79 20.78 20.62			
Mean	(1 10)	18.42	0.17	0.90		20.66	0.19	0.94

for containing a partial cDNA for human factor VIII, pF8TM. A series of transfections was set up using a decreasing amount of the plasmid (40, 4, 0.5, and 0.1 µg). Twenty-four hours post-transfection, total DNA was purified from each flask of cells. β-Actin gene quantity was chosen as a value for normalization of genomic DNA concentration from each sample. In this experiment, β-actin gene content should remain constant relative to total genomic DNA. Figure 3 shows the result of the β-actin DNA measurement (100 ng total DNA determined by ultraviolet spectroscopy) of each sample. Each sample was analyzed in triplicate and the mean β-actin C_T values of the triplicates were plotted (error bars represent one standard deviation). The highest difference

between any two sample means was 0.95 C_T. Ten nanograms of total DNA of each sample were also examined for β-actin. The results again showed that very similar amounts of genomic DNA were present; the maximum mean β-actin C_T value difference was 1.0. As Figure 3 shows, the rate of β-actin C_T change between the 100 and 10-ng samples was similar (slope values range between 3.56 and -3.45). This verifies again that the method of sample preparation yields samples of identical PCR integrity (i.e., no sample contained an excessive amount of a PCR inhibitor). However, these results indicate that each sample contained slight differences in the actual amount of genomic DNA analyzed. Determination of actual genomic DNA concentration was accomplished

REAL TIME QUANTITATIVE PCR

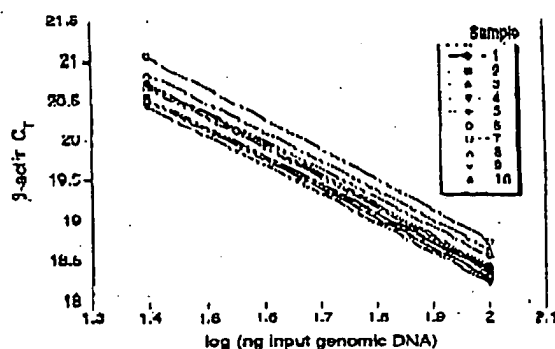


Figure 2 Sample preparation purity. The replicate samples shown in Table 1 were also amplified in triplicate using 25 ng of each DNA sample. The figure shows the input DNA concentration (100 and 25 ng) vs. C_t . In the figure, the 100 and 25 ng points for each sample are connected by a line.

by plotting the mean β -actin C_t value obtained for each 100-ng sample on a β -actin standard curve (shown in Fig. 4C). The actual genomic DNA concentration of each sample, a , was obtained by extrapolation to the x-axis.

Figure 4A shows the measured (i.e., non-normalized) quantities of factor VIII plasmid DNA (pF8TM) from each of the four transient cell transfections. Each reaction contained 100 ng of total sample DNA (as determined by UV spectroscopy). Each sample was analyzed in triplicate

PCR amplifications. As shown, pF8TM purified from the 293 cells decreases (mean C_t values increase) with decreasing amounts of plasmid distributed. The mean C_t values obtained for pF8TM in Figure 4A were plotted on a standard curve comprised of serially diluted pF8TM, shown in Figure 4B. The quantity of pF8TM, b , found in each of the four transfections was determined by extrapolation to the x-axis of the standard curve in Figure 4B. These uncorrected values, b , for pF8TM were normalized to determine the actual amount of pF8TM found per 100 ng of genomic DNA by using the equation:

$$\frac{b \times 100 \text{ ng}}{a} = \text{actual pF8TM copies per 100 ng of genomic DNA}$$

where a = actual genomic DNA in a sample and b = pF8TM copies from the standard curve. The normalized quantity of pF8TM per 100 ng of genomic DNA for each of the four transfections is shown in Figure 4D. These results show that the quantity of factor VIII plasmid associated with the 293 cells, 24 hr after transfection, decreases with decreasing plasmid concentration used in the transfection. The quantity of pF8TM associated with 293 cells, after transfection with 40 μ g of plasmid, was 35 pg per 100 ng genomic DNA. This results in ~520 plasmid copies per cell.

DISCUSSION

We have described a new method for quantitating gene copy numbers using real-time analysis of PCR amplifications. Real-time PCR is compatible with either of the two PCR (RT-PCR) approaches: (1) quantitative competitive where an internal competitor for each target sequence is used for normalization (data not shown) or (2) quantitative comparative PCR using a normalization gene contained within the sample (i.e., β -actin) or a "housekeeping" gene for RT-PCR. If equal amounts of nucleic acid are analyzed for each sample and if the amplification efficiency before quantitative analysis is identical for each sample, the internal control (normalization gene or competitor) should give equal signals for all samples.

The real-time PCR method offers several advantages over the other two methods currently employed (see the Introduction). First, the real-time PCR method is performed in a closed-tube system and requires no post-PCR manipulation

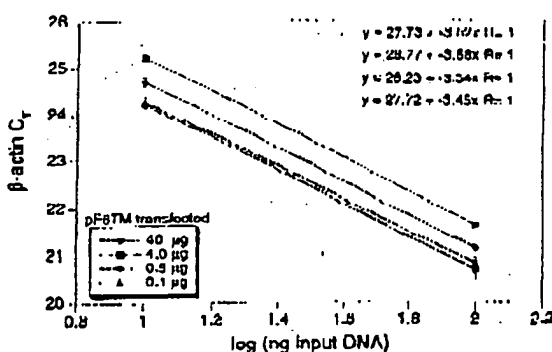


Figure 3 Analysis of transfected cell DNA quantity and purity. The DNA preparations of the four 293 cell transfections (40, 4, 0.5, and 0.1 μ g of pF8TM) were analyzed for the β -actin gene. 100 and 10 ng (determined by ultraviolet spectroscopy) of each sample were amplified in triplicate. For each amount of pF8TM that was transfected, the β -actin C_t values are plotted versus the total input DNA concentration.

HLID ET AL.

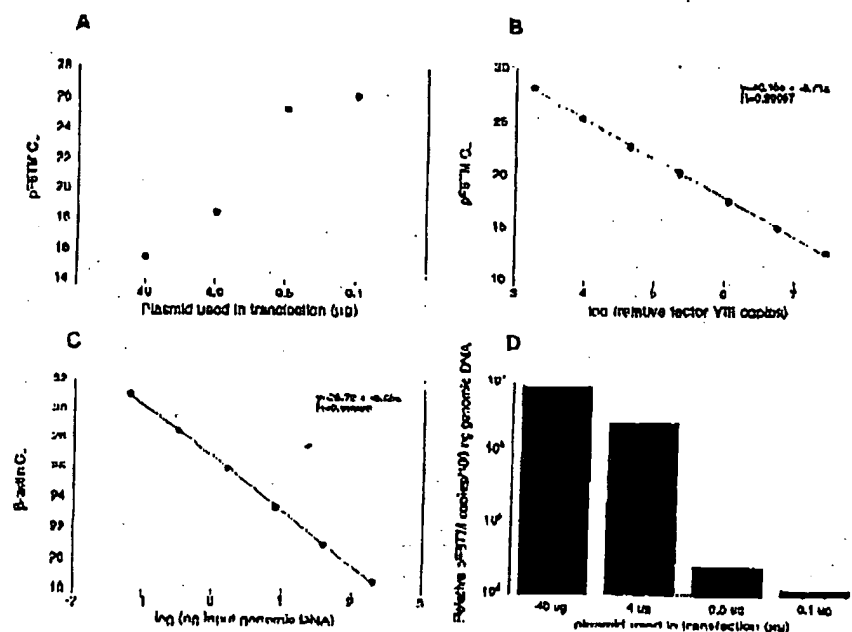


Figure 4 Quantitative analysis of pF8TM in transfected cells. (A) Amount of plasmid DNA used for the transfection plotted against the mean C_T value determined for pF8TM remaining 24 hr after transfection. (B,C) Standard curves of pF8TM and β -actin, respectively. pF8TM DNA (B) and genomic DNA (C) were diluted serially 1:5 before amplification with the appropriate primers. The β -actin standard curve was used to normalize the results of A to 100 ng of genomic DNA. (D) The amount of pF8TM present per 100 ng of genomic DNA.

of sample. Therefore, the potential for PCR contamination in the laboratory is reduced because amplified products can be analyzed and disposed of without opening the reaction tubes. Second, this method supports the use of a normalization gene (i.e., β -actin) for quantitative PCR or house-keeping genes for quantitative RT-PCR controls. Analysis is performed in real time during the log phase of product accumulation. Analysis during log phase permits many different genes (over a wide input target range) to be analyzed simultaneously, without concern of reaching reaction plateau at different cycles. This will make multi-gene analysis assays much easier to develop, because individual internal competitors will not be needed for each gene under analysis. Third, sample throughput will increase dramatically with the new method because there is no post-PCR processing time. Additionally, working in a 96-well format is highly compatible with automation technology.

The real-time PCR method is highly reproducible. Replicate amplifications can be analyzed

for each sample minimizing potential error. The system allows for a very large assay dynamic range (approaching 1,000,000-fold starting target). Using a standard curve for the target of interest, relative copy number values can be determined for any unknown sample. Fluorescent threshold values, C_T , correlate linearly with relative DNA copy numbers. Real time quantitative RT-PCR methodology (Gibson et al., this issue) has also been developed. Finally, real time quantitative PCR methodology can be used to develop high-throughput screening assays for a variety of applications [quantitative gene expression (RT-PCR), gene copy assays (Her2, HIV, etc.), genotyping (knockout mouse analysis), and immunopCR].

Real-time PCR may also be performed using intercalating dyes (Higuchi et al. 1992) such as ethidium bromide. The fluorogenic probe method offers a major advantage over intercalating dyes—greater specificity (i.e., primer dimers and nonspecific PCR products are not detected).

METHODS

Generation of a Plasmid Containing a Partial cDNA for Human Factor VIII

Total RNA was harvested (RNAzol B from Tel Test, Inc., Friendswood, TX) from cells transfected with a factor VIII expression vector, pCIS2.8(251) (Eaton et al. 1986; Gorman et al. 1990). A factor VIII partial cDNA sequence was generated by RT-PCR [GeneAmp EZ RT-PCR Kit (part N808-0179, PE Applied Biosystems, Foster City, CA)] using the PCR primers F8for and F8rev (primer sequences are shown below). The amplicon was reamplified using modified F8for and F8rev primers (appended with *Hind*III and *Hind*III restriction site sequences at the 5' end) and cloned into pE8M-32 (Promega Corp., Madison, WI). The resulting clone, pF8TM, was used for transient transfection of 293 cells.

Amplification of Target DNA and Detection of Amplicon Factor VIII Plasmid DNA

(pF8TM) was amplified with the primers F8for 5'-GAGGCTTCCCAAGAGTGAAGTTC-3' and F8rev 5'-AAACCTTCAAGCTTGGATCGTACG-3'. The reaction produced a 422-bp PCR product. The forward primer was designed to recognize a unique sequence found in the 5' untranslated region of the parent pCIS2.8(251) plasmid and therefore does not recognize and amplify the human factor VIII gene. Primers were chosen with the assistance of the computer program Oligo 4.0 (National Biosciences, Inc., Plymouth, MN). The human β -actin gene was amplified with the primers β -actin forward primer 5'-TCACCCACACCTTGTCCCATCTACCA-3' and β -actin reverse primer 5'-CAGCGGAACCCCTTCATTGCAATGG-3'. The reaction produced a 295-bp PCR product.

Amplification reactions (50 μ l) contained a DNA sample, 10 \times PCR Buffer II (5 μ l), 200 μ M dATP, dCTP, dGTP, and 400 μ M dUTP, 4 mM MgCl₂, 1.25 Units AmpliTaq DNA polymerase, 0.5 unit Amplifase uracil N-glycosylase (UNG), 50 pmole of each factor VIII primer, and 15 pmole of each β -actin primer. The reactions also contained one of the following detection probes (100 nM each): F8probe 5'-(FAM)AGCTCTTCCACCTTCTTCTTCTTCTTCTTGCCTT(TAMRA)p 3' and β -actin probe 5'-(FAM)ATGCCCX(TAMRA)CCCCCATGCCATCp-3' where p indicates phosphorylation and X indicates a linker arm nucleotide. Reaction tubes were MicroAmp Optical Tubes (part number N801 0933, Perkin Elmer) that were frosted (at Perkin Elmer) to prevent light from reflecting. Tube caps were similar to MicroAmp Caps but specially designed to prevent light scattering. All of the PCR consumables were supplied by PE Applied Biosystems (Foster City, CA) except the factor VIII primers, which were synthesized at Genentech, Inc. (South San Francisco, CA). Probes were designed using the Oligo 4.0 software, following guidelines suggested in the Model 7700 Sequence Detector Instrument manual. Briefly, probe T_m should be at least 5°C higher than the annealing temperature used during thermal cycling; primers should not form stable duplexes with the probe.

The thermal cycling conditions included 2 min at 50°C and 10 min at 95°C. Thermal cycling proceeded with

REAL TIME QUANTITATIVE PCR

reactions were performed in the Model 7700 Sequence Detector (PE Applied Biosystems), which contains a GeneAmp PCR System 9600. Reaction conditions were programmed on a Power Macintosh 7100 (Apple Computer, Santa Clara, CA) linked directly to the Model 7700 Sequence Detector. Analysis of data was also performed on the Macintosh computer. Collection and analysis software was developed at PE Applied Biosystems.

Transfection of Cells with Factor VIII Construct

Four T175 flasks of 293 cells (ATCC CRL 1573), a human fetal kidney suspension cell line, were grown to 80% confluency and transfected pF8TM. Cells were grown in the following media: 50% HAM'S F12 without GHT, 50% low glucose Dulbecco's modified Eagle medium (DMEM) without glycine with sodium bicarbonate, 10% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin-streptomycin. The media was changed 30 min before the transfection. pF8TM DNA amounts of 40, 4, 0.5, and 0.1 μ g were added to 1.5 ml of a solution containing 0.125 M CaCl₂ and 1 \times HBPS. The four mixtures were left at room temperature for 10 min and then added dropwise to the cells. The flasks were incubated at 37°C and 5% CO₂ for 24 hr, washed with PBS, and resuspended in PBS. The transfected cells were divided into aliquots and DNA was extracted immediately using the QIAamp Blood Kit (Qiagen, Chatsworth, CA). DNA was eluted into 200 μ l of 20 mM Tris-HCl at pH 8.0.

ACKNOWLEDGMENTS

We thank Genentech's DNA Synthesis Group for primer synthesis and Genentech's Graphics Group for assistance with the figures.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bassler, H.A., S.J. Flood, K.J. Livak, J. Marmaro, R. Kohn, and C.A. Batt. 1995. Use of a fluorogenic probe in a PCR-based assay for the detection of *Listeria monocytogenes*. *App. Environ. Microbiol.* 61: 3724-3728.
- Buckler-Andre, M. 1991. Quantitative evaluation of mRNA levels. *Math. Mol. Cell. Biol.* 2: 189-201.
- Clementi, M., S. Menzo, P. Bugnarelli, A. Manzini, A. Valenza, and P.E. Varaldo. 1993. Quantitative PCR and RT-PCR in virology. [Review]. *PCR Methods Applic.* 2: 191-196.
- Connor, R.J., H. Mohl, Y. Cao, and D.D. Ho. 1993. Increased viral burden and cytopathicity correlate temporally with CD4⁺ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 67: 1772-1777.
- Eaton, D.L., W.J. Wood, D. Eaton, P.E. Hass, P.

HFID LI AL

- Venar, and C. Gornum. 1986. Construction and characterization of an active factor VIII variant lacking the central one third of the molecule. *Biochemistry* 25: 8343-8347.
- Fasco, M.J., C.P. Treanor, S. Spivack, H.L. Wigge, and L.S. Kaminsky. 1995. Quantitative RNA-polymerase chain reaction-DNA analysis by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* 224: 140-147.
- Ferre, R. 1992. Quantitative or semi-quantitative PCR: Reality versus myth. *PCR Methods Applic.* 2: 1-9.
- Furtado, M.R., L.A. Kingsley, and S.M. Wollinsky. 1995. Changes in the viral mRNA expression pattern correlate with a rapid rate of CD4+ T-cell number decline in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 69: 2097-2100.
- Gibson, U.E.M., C.A. Heid, and P.M. Williams. 1996. A novel method for real time quantitative competitive RT-PCR. *Genome Res.* (this issue).
- Gorman, C.M., D.R. Gies, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Prot. Engin. Tech.* 2: 3-10.
- Higuchi, R., G. Dollinger, P.S. Walsh, and R. Griffith. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biochemistry* 10: 413-417.
- Holland, P.M., R.D. Abramson, R. Watson, and D.J. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* 88: 7276-7280.
- Huang, S.K., H.Q. Xiao, T.J. Klein, G. Paciotti, H.G. Marsh, L.M. Lichtenstein, and M.C. Liu. 1995a. IL-13 expression at the sites of allergen challenge in patients with asthma. *J. Immun.* 155: 2688-2694.
- Huang, S.K., M. Yi, E. Palmer, and D.G. Marsh. 1995b. A dominant T cell receptor beta-chain in response to a short ragweed allergen, Amb a 5. *J. Immun.* 154: 6157-6162.
- Kellogg, D.E., J.J. Shinsky, and S. Kowk. 1990. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. *Anal. Biochem.* 189: 202-208.
- Lee, J.-G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761-3766.
- Livak, K.J., S.J. Flood, J. Matraro, W. Gust, and K. Deetz. 1995a. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applic.* 4: 357-362.
- Livak, K.J., J. Matraro, and J.A. Todd. 1995b. Towards fully automated genome-wide polymorphism screening [Letter]. *Nature Genet.* 9: 341-342.
- Mulder, J., N. McKinney, C. Christopherson, J. Sullsky, L. Greenfield, and S. Kwok. 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: Application to acute retroviral infection. *J. Clin. Microbiol.* 32: 292-300.
- Pang, S., Y. Koyanagi, S. Miller, C. Wiloy, H.V. Vinters, and L.S. Chen. 1990. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343: 85-89.
- Platak, M.J., K.C. Luk, B. Williams, and J.D. Lifson. 1993a. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *BioTechniques* 14: 70-81.
- Platak, M.J., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hann, G.M. Shaw, and J.D. Lifson. 1993b. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR [see Comments]. *Science* 259: 1749-1754.
- Prud'homme, G.J., D.H. Kono, and A.N. Theofilopoulos. 1995. Quantitative polymerase chain reaction analysis reveals marked overexpression of interleukin-1 beta, interleukin-1 and interferon-gamma mRNA in the lymph nodes of lupus-prone mice. *Mol. Immunol.* 32: 495-503.
- Racymackers, L. 1995. A commentary on the practical applications of competitive PCR. *Genome Res.* 5: 91-94.
- Sharp, P.A., A.J. Berk, and S.M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* 65: 750-768.
- Slamon, D.J., G.M. Clark, S.C. Wong, W.J. Levin, A. Ullrich, and W.L. McGuire. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177-182.
- Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.
- Tan, X., X. Sun, C.F. Gonzalez, and W. Hsueh. 1994. PAF and TNF increase the precursor of NF-kappa B p50 mRNA in mouse intestine: Quantitative analysis by competitive PCR. *Biochim. Biophys. Acta* 1215: 157-162.
- Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci.* 77: 5201-5205.
- Williams, S., C. Schwer, A. Krishnasao, C. Held, B. Karger, and P.M. Williams. 1996. Quantitative competitive PCR: Analysis of amplified products of the HIV-1 gag gene by capillary electrophoresis with laser induced fluorescence detection. *Anal. Biochem.* (in press).

Received June 3, 1996; accepted in revised form July 29, 1996.

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA*†, TODD A. SWANSON*, JAMES W. WELSH*, MARGARET A. ROY‡, DAVID A. LAWRENCE*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE||, ROBERT L. COHEN*, MONA F. MELHEM**, GENE G. FINLEY**, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,‡‡, AND ARNOLD J. LEVINE§

Departments of *Molecular Oncology, ‡Molecular Biology, §Scientific Computing, and ¶Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; **University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 β (GSK-3 β) resulting in an increase in β -catenin levels. Stabilized β -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β -catenin levels (9). APC is phosphorylated by GSK-3 β , binds to β -catenin, and facilitates its degradation. Mutations in either APC or β -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- β superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twist* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0
PNAS is available online at www.pnas.org.

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

cDNA was synthesized from 2 μ g of poly(A)⁺ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μ g of poly(A)⁺ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

In Situ Hybridization. ³³P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNazol.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{(-\Delta Ct)}$ where ΔCt represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The δ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of $\approx 40,000$ (M_r , 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of $\approx 27,000$ (M_r , 27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

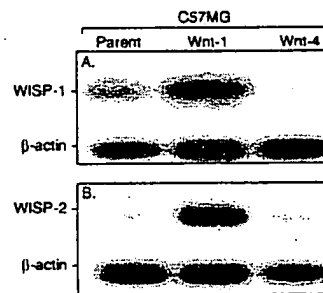


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)⁺ RNA (2 μ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human β -actin probe.

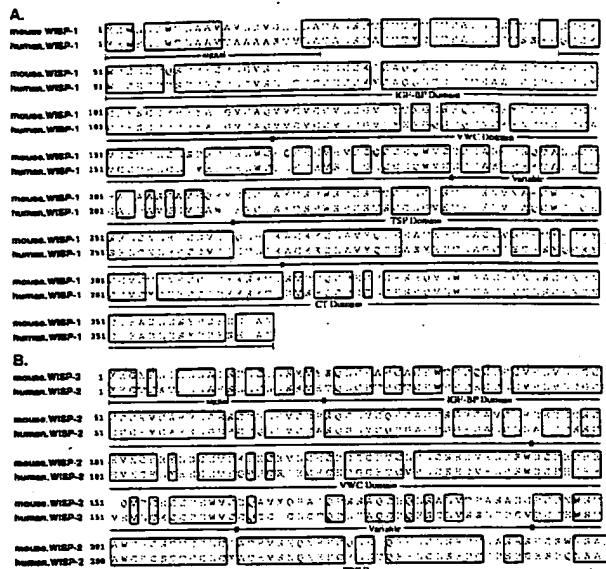


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

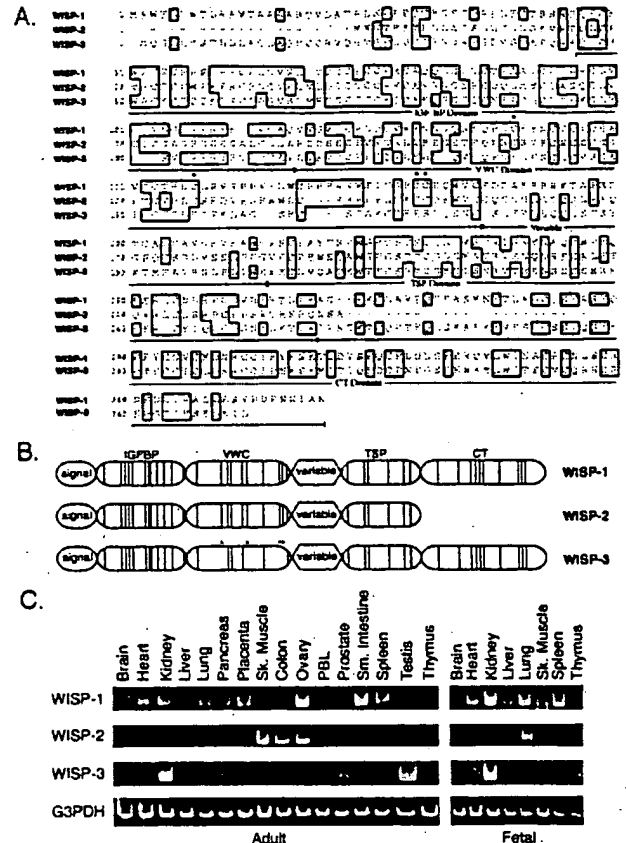


FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSxCSSxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

In Situ Localization of *WISP-1* and *WISP-2*. Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

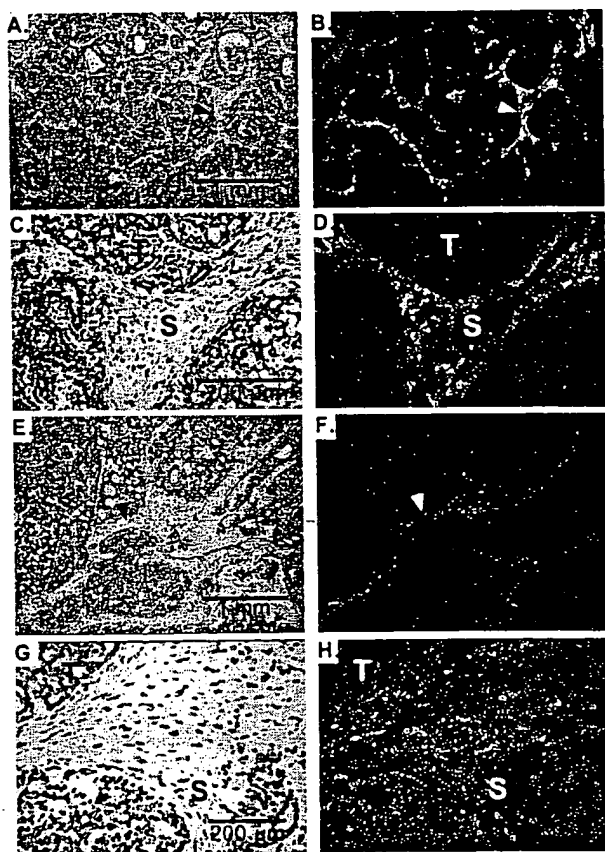


FIG. 4. (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (G and H).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

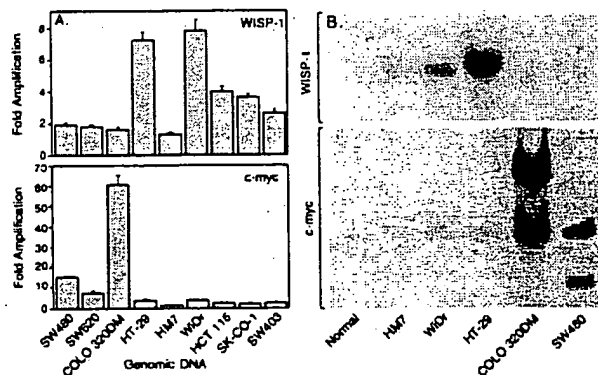


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10 μ g) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

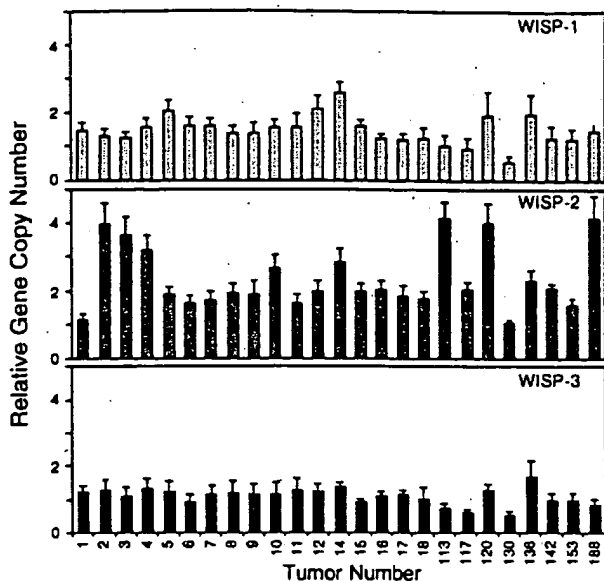


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

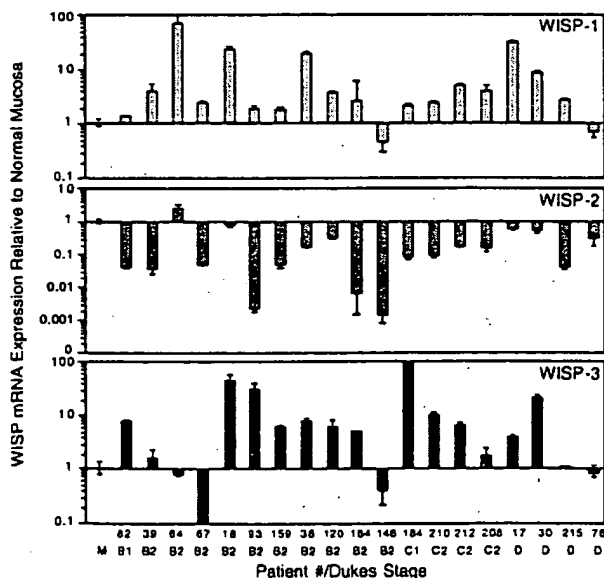


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- β , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_v\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- β 1, which is the stimulus for stromal proliferation (34). TGF- β 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISPs*. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISPs* as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISPs* in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

1. Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286–3305.
2. Dale, T. C. (1998) *Biochem. J.* 329, 209–223.
3. Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99–109.
4. van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233–240.
5. Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619–625.
6. Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182–187.
7. Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391–399.
8. Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248–1256.
9. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046–3050.
10. He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509–1512.
11. Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025–6030.
12. Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001–1009.
13. Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278–6286.
14. Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349–1358.
15. Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289–296.
16. Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131–6141.
17. Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171–179.
18. Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326–1334.
19. Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355–6360.
20. Martinier, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729–2732.
21. Bork, P. (1993) *FEBS Lett.* 327, 125–130.
22. Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981–12986.
23. Joliot, V., Martinier, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10–21.
24. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514–19527.
25. Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852–2855.
26. Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195–205.
27. Martinier, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529–2534.
28. Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell Genet.* 57, 109–111.
29. Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88–94.
30. Garte, S. J. (1993) *Crit. Rev. Oncog.* 4, 435–449.
31. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268–1272.
32. Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269–291.
33. Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090–3096.
34. Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153–161.
35. Wernert, N. (1997) *Virchows Arch.* 430, 433–443.
36. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., et al. (1994) *Cancer Res.* 54, 4257–4260.
37. Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187–194.
38. Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., et al. (1998) *EMBO J.* 17, 3052–3065.
39. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787–1790.
40. Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169–176.

methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described^{20,21}. Briefly, after antigen pulsing (30 µg ml⁻¹ TTCF) with tetrapeptides (1–2 mg ml⁻¹), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with 1 µCi of ³H-thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating 200 µg TTCF with 0.25 µg pig kidney legumain in 500 µl 50 mM citrate buffer, pH 5.5, for 1 h at 37 °C. **Glycopeptide digestions.** The peptides HIDNEED, HIDN(N-glucosamine) EEDI and HIDNESDI, which are based on the TTCF sequence, and QQQLHFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQLHFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography¹¹. Glycopeptides corresponding to residues 622–642 and 421–452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at 30 °C with 5–50 mU ml⁻¹ pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of 10 mg ml⁻¹ α-cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 3 November 1998.

- Chen, J. M. et al. Cloning, isolation, and characterisation of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090–8098 (1997).
- Kembhavi, A. A., Buttle, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 303, 208–213 (1993).
- Dalton, J. P., Hala Jamriska, L. & Bridley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* 111, 575–580 (1995).
- Bennett, K. et al. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* 22, 1519–1524 (1992).
- Riese, R. J. et al. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357–366 (1996).
- Rodríguez, G. M. & Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* 149, 2894–2898 (1992).
- Hewitt, E. W. et al. Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* 159, 4693–4699 (1997).
- Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821–850 (1997).
- Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* 10, 93–102 (1998).
- Fineschi, B. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* 22, 377–382 (1997).
- Lu, J. & van Halbeek, H. Complete ¹H and ¹³C resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* 296, 1–21 (1996).
- Fearon, D. T. & Lockley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* 272, 50–54 (1996).
- Medzhitov, R. & Janeway, C. A. J. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91, 295–298 (1997).
- Wyatt, R. et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705–711 (1998).
- Botarelli, P. et al. N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* 147, 3128–3132 (1991).
- Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* 144, 4101–4109 (1990).
- Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* 80, 535–559 (1981).
- Makoff, A. J., Ballantine, S. P., Smallwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* 7, 1043–1046 (1989).
- Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
- Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* 314, 537–539 (1985).
- Pond, L. & Watts, C. Characterization of transport of newly assembled, T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* 159, 543–553 (1997).

Acknowledgements. We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti[†], Scot A. Marsters[†], David A. Lawrence[†], Margaret Roy^{*}, Frank C. Kischkel^{*}, Patrick Dowd^{*}, Arthur Huang^{*}, Christopher J. Donahue^{*}, Steven W. Sherwood^{*}, Daryl T. Baldwin^{*}, Paul J. Godowski^{*}, William I. Wood^{*}, Austin L. Gurney^{*}, Kenneth J. Hillan^{*}, Robert L. Cohen^{*}, Audrey D. Goddard^{*}, David Botstein[‡] & Avi Ashkenazi^{*}

^{*} Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

[‡] Department of Genetics, Stanford University, Stanford, California 94305, USA

[†] These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells¹. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (DcR3), that binds to FasL and inhibits FasL-induced apoptosis. The DcR3 gene was amplified in about half of 35 primary lung and colon tumours studied, and DcR3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily². Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (DcR3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)³, DcR3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-associated, molecule. We expressed a recombinant, histidine-tagged form of DcR3 in mammalian cells; DcR3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). DcR3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of DcR3 and OPG are conserved; however, the carboxy-terminal portion of DcR3 is 101 residues shorter.

We analysed expression of DcR3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high DcR3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of DcR3, we generated a recombinant, Fc-tagged DcR3 protein. We tested binding of DcR3-Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). DcR3-Fc showed a significant increase in binding to cells transfected with FasL⁴ (Fig. 2a), but not to cells transfected with TNF⁵, Apo2L/TRAIL^{6,7}, Apo3L/TWEAK^{8,9}, or OPGL/TRACE/

RANKL¹⁰⁻¹² (data not shown). DcR3-Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3-Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3-Fc and Fas-Fc bound to soluble FasL with a comparable affinity ($K_d = 0.8 \pm 0.2$ and 1.1 ± 0.1 nM, respectively; Fig. 2e), and that DcR3-Fc could block nearly all of the binding of soluble FasL to Fas-Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3-Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3-Fc and Fas-Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at $\sim 0.1 \mu\text{g ml}^{-1}$. Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3-Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process¹. Consistent with previous results¹³, activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas-Fc blocked this effect substantially (Fig. 3c); DcR3-Fc blocked the

induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes^{1,14-16}. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3-Fc and Fas-Fc each reduced killing of target cells from $\sim 65\%$ to $\sim 30\%$, with half-maximal inhibition at $\sim 1 \mu\text{g ml}^{-1}$; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3-Fc and Fas-Fc concentrations were required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL¹⁷.

Given the role of immune-cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain

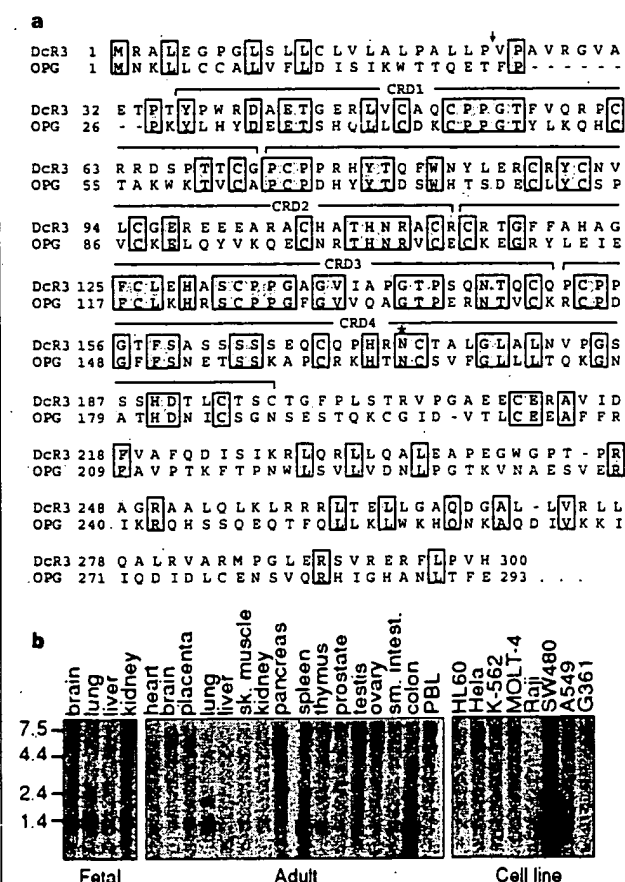


Figure 1 Primary structure and expression of human DcR3. **a**, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1-4), and the N-linked glycosylation site (asterisk) are shown. **b**, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)⁺ RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.

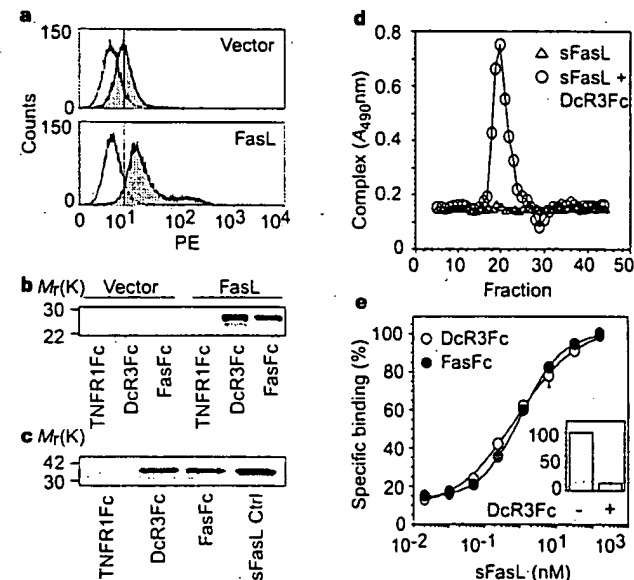


Figure 2 Interaction of DcR3 with FasL. **a**, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3-Fc (solid line, shaded area), TNFR1-Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ($P < 0.001$) between the binding of DcR3-Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. **b**, 293 cells were transfected as in **a** and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. **c**, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1-Fc, DcR3-Fc or Fas-Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. **d**, Flag-tagged sFasL was incubated with DcR3-Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3-Fc and sFasL-Flag. **e**, Equilibrium binding of DcR3-Fc or Fas-Fc to sFasL-Flag. Inset, competition of DcR3-Fc with Fas-Fc for binding to sFasL-Flag.

reaction (PCR)¹⁸ in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this,

we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM218xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epicentre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG^{2,19}.

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas²⁰. A second mechanism involves proteolytic shedding of FasL from the cell surface¹⁷. DcR3 competes with Fas for

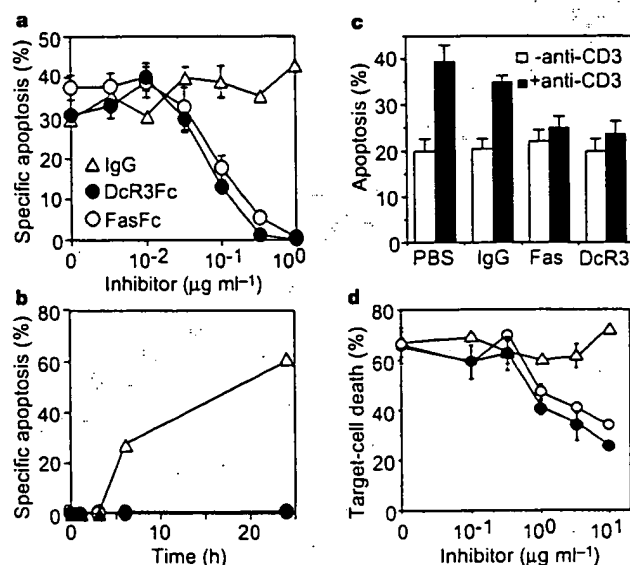


Figure 3 Inhibition of FasL activity by DcR3. **a**, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml⁻¹) oligomerized with anti-Flag antibody (0.1 μg ml⁻¹) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). **b**, Jurkat cells were incubated with sFasL-Flag plus anti-Flag antibody as in **a**, in presence of 1 μg ml⁻¹ DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. **c**, Peripheral blood T cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml⁻¹). After 16 h, apoptosis of CD4⁺ cells was determined (mean ± s.e.m. of results from five donors). **d**, Peripheral blood natural killer cells were incubated with ⁵¹Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of ⁵¹Cr (mean ± s.d. for two donors, each in triplicate).

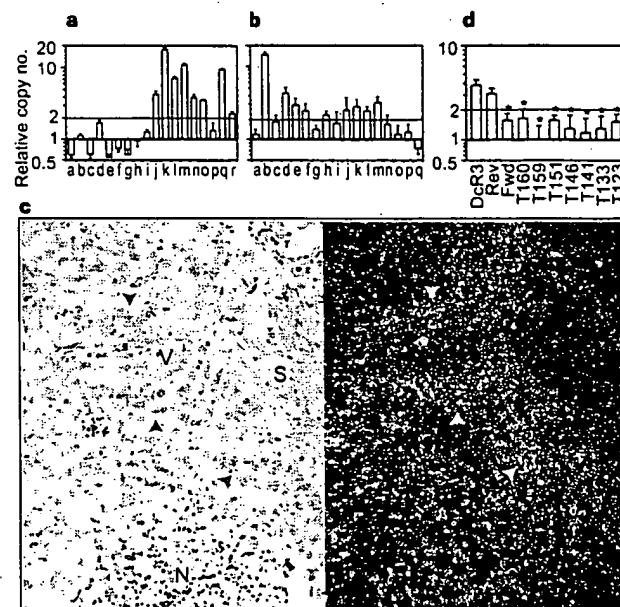


Figure 4 Genomic amplification of DcR3 in tumours. **a**, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, r), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (i), and one bronchial adenocarcinoma (l). The data are means ± s.d. of 2 experiments done in duplicate. **b**, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. **c**, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. **d**, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (**b**). Data are from two experiments done in duplicate. Asterisk indicates $P < 0.01$ for a Student's *t*-test comparing each marker with DcR3.

FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described²¹. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L²². Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG³, which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L¹⁹. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response². Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours.

Methods

Isolation of DcR3 cDNA. Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in LifeseqTM (Incyte Pharmaceuticals; accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

Fc-fusion proteins (immunoadhesins). The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described²³.

Fluorescence-activated cell sorting (FACS) analysis. We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRK5 vector or pRK5 encoding full-length human FasL⁴ (2 µg), together with pRK5 encoding CrmA (2 µg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3-Fc or TNFR1-Fc and then with phycoerythrin-conjugated streptavidin (GibcoBRL), and were assayed by FACS. The data were analysed by Kolmogorov-Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3-Fc; as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

Immunoprecipitation. Human 293 cells were transfected as above, and metabolically labelled with [³⁵S]cysteine and [³⁵S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 µM), the medium was immunoprecipitated with DcR3-Fc, Fas-Fc or TNFR1-Fc (5 µg), followed by protein A-Sepharose (Repligen). The precipitates were resolved by SDS-PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 µg) (Alexis) was incubated with each Fc-fusion protein (1 µg), precipitated with protein A-Sepharose, resolved by SDS-PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

Analysis of complex formation. Flag-tagged soluble FasL (25 µg) was incubated with buffer or with DcR3-Fc (40 µg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3-Fc-FasL complex in each fraction was analysed by placing 100 µl aliquots into microtitre wells precoated with anti-human IgG (Boehringer) to capture DcR3-Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin-horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3-Fc homodimers to two soluble FasL homotrimers.

Equilibrium binding analysis. Microtitre wells were coated with anti-human

IgG, blocked with 2% BSA in PBS. DcR3-Fc or Fas-Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas-Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3-Fc.

T-cell AICD. CD3⁺ lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotec), stimulated with phytohaemagglutinin (PHA; 2 µg ml⁻¹) for 24 h, and cultured in the presence of interleukin-2 (100 U ml⁻¹) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4⁺ cells²⁴.

Natural killer cell activity. Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotec), and incubated for 16 h with ⁵¹Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3-Fc, Fas-Fc or human IgG1. Target-cell death was determined by release of ⁵¹Cr in effector-target cocultures relative to release of ⁵¹Cr by detergent lysis of equal numbers of Jurkat cells.

Gene-amplification analysis. Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorometry. Amplification was determined by quantitative PCR¹⁸ using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene; alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218xe7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5'-CTTCTTCGCGCAGCTG-3' and 5'-ATCACGCCGGCACCAG-3' and the fluorogenic probe sequence was 5'-(FAM-ACACGATGCGTGCTCCAAGCAG Aap-(TAMARA), where FAM is 5'-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula 2^(ΔCT), where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

- Nagata, S. Apoptosis by death factor. *Cell* 88, 355-365 (1997).
- Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* 76, 959-962 (1994).
- Simonet, W. S. et al. Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 89, 309-319 (1997).
- Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* 75, 1169-1178 (1993).
- Pennica, D. et al. Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* 312, 724-729 (1984).
- Pitti, R. M. et al. Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* 271, 12687-12690 (1996).
- Wiley, S. R. et al. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673-682 (1995).
- Marsters, S. A. et al. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
- Chicheportiche, Y. et al. TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* 272, 32401-32410 (1997).
- Wong, B. R. et al. TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* 272, 25190-25194 (1997).
- Anderson, D. M. et al. A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* 390, 175-179 (1997).
- Lacey, D. L. et al. Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* 93, 165-176 (1998).
- Dhein, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo1/Fas/CD95. *Nature* 373, 438-441 (1995).
- Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* 181, 1235-1238 (1995).
- Medvedev, A. E. et al. Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* 9, 394-404 (1997).
- Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* 90, 13-18 (1997).
- Tanaka, M., Imai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* 4, 31-36 (1998).
- Gelmini, S. et al. Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* 43, 752-758 (1997).
- Emery, J. G. et al. Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* 273, 14363-14367 (1998).
- Wallach, D. Placing death under control. *Nature* 388, 123-125 (1997).
- Collotta, F. et al. Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. *Science* 261, 472-475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* **281**, 1305–1308 (1998).
23. Ashkenazi, A. & Chomow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195–200 (1997).
24. Masters, S. *et al.* Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* **6**, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the DcR3 cDNA sequence is AF104419.

Crystal structure of the ATP-binding subunit of an ABC transporter

Li-Wei Hung*, Iris Xiaoyan Wang†, Kishiko Nikaido†, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim*‡

* E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes¹. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters². Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains¹. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*^{3–8} is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, HisQMP₂, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins⁹, is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM⁶. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis⁵, the requirement for both subunits to be present for activity⁸, and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer³. HisP has been purified and characterized in an active soluble form³ which can be reconstituted into a fully active membrane-bound complex⁸.

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded β -sheet ($\beta 3$ and $\beta 8$ – $\beta 12$) spans both arms of the L, with a domain of α - plus β -type structure ($\beta 1$, $\beta 2$, $\beta 4$ – $\beta 7$, $\alpha 1$ and $\alpha 2$) on one side (within arm I) and a domain of mostly α -helices ($\alpha 3$ – $\alpha 9$) on the

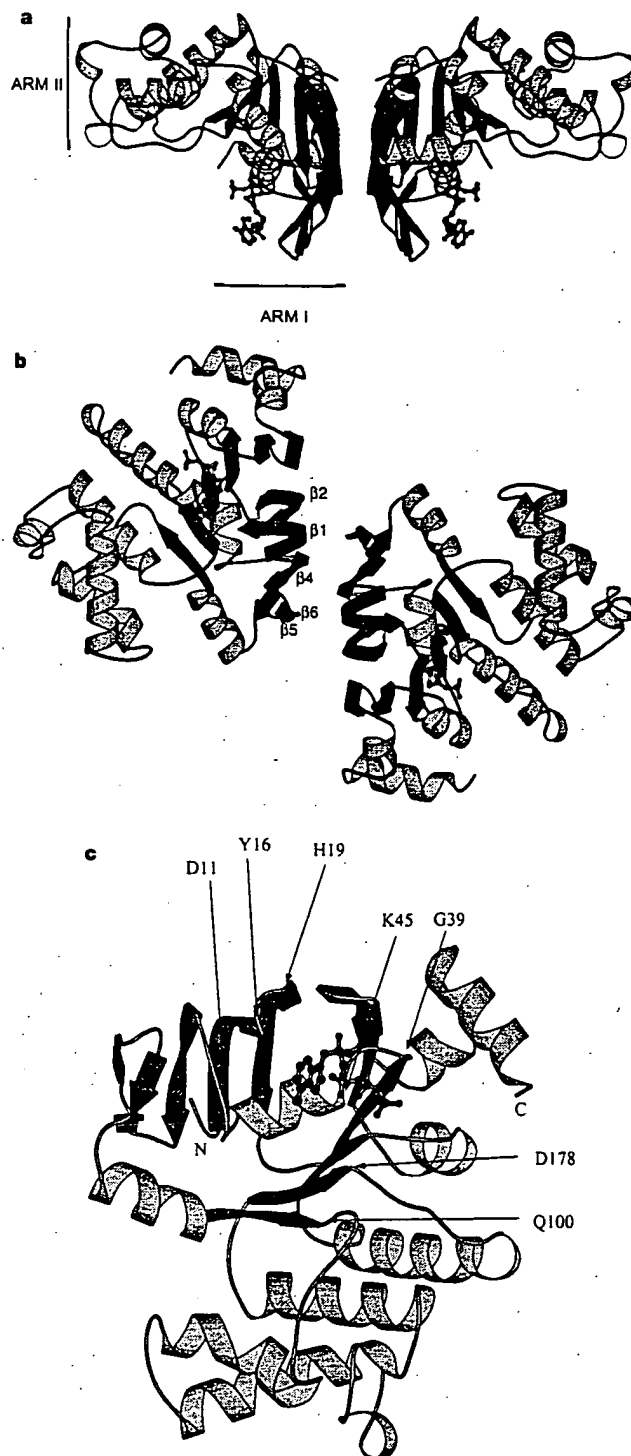


Figure 1 Crystal structure of HisP. **a**, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane. α -Helices are shown in orange and β -sheets in green. **b**, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in **a**. The β -strands at the dimer interface are labelled. **c**, View of one monomer from the bottom of arm I, as shown in **a**, towards arm II, showing the ATP-binding pocket. **a**–**c**, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in **c**. These figures were prepared with MOLSCRIPT²⁸. N, amino terminus; C, C terminus.

NOVEL APPROACH TO QUANTITATIVE POLYMERASE CHAIN REACTION USING REAL-TIME DETECTION: APPLICATION TO THE DETECTION OF GENE AMPLIFICATION IN BREAST CANCER

Ivan BIÈCHE^{1,2}, Martine OLIVI¹, Marie-Hélène CHAMPÈME², Dominique VIDAUD¹, Rosette LIDEREAU² and Michel VIDAUD^{1*}

¹Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, Paris, France

²Laboratoire d'Oncogénétique, Centre René Huguenin, St-Cloud, France

Gene amplification is a common event in the progression of human cancers, and amplified oncogenes have been shown to have diagnostic, prognostic and therapeutic relevance. A kinetic quantitative polymerase-chain-reaction (PCR) method, based on fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real-time, was used to quantify gene amplification in tumor DNA. Reactions are characterized by the point during cycling when PCR amplification is still in the exponential phase, rather than the amount of PCR product accumulated after a fixed number of cycles. None of the reaction components is limited during the exponential phase, meaning that values are highly reproducible in reactions starting with the same copy number. This greatly improves the precision of DNA quantification. Moreover, real-time PCR does not require post-PCR sample handling, thereby preventing potential PCR-product carry-over contamination; it possesses a wide dynamic range of quantification and results in much faster and higher sample throughput. The real-time PCR method, was used to develop and validate a simple and rapid assay for the detection and quantification of the 3 most frequently amplified genes (*myc*, *ccnd1* and *erbB2*) in breast tumors. Extra copies of *myc*, *ccnd1* and *erbB2* were observed in 10, 23 and 15%, respectively, of 108 breast-tumor DNA; the largest observed numbers of gene copies were 4.6, 18.6 and 15.1, respectively. These results correlated well with those of Southern blotting. The use of this new semi-automated technique will make molecular analysis of human cancers simpler and more reliable, and should find broad applications in clinical and research settings. *Int. J. Cancer* 78:661–666, 1998.

© 1998 Wiley-Liss, Inc.

Gene amplification plays an important role in the pathogenesis of various solid tumors, including breast cancer, probably because over-expression of the amplified target genes confers a selective advantage. The first technique used to detect genomic amplification was cytogenetic analysis. Amplification of several chromosome regions, visualized either as extrachromosomal double minutes (dmns) or as integrated homogeneously staining regions (HSRs), are among the main visible cytogenetic abnormalities in breast tumors. Other techniques such as comparative genomic hybridization (CGH) (Kallioniemi *et al.*, 1994) have also been used in broad searches for regions of increased DNA copy numbers in tumor cells, and have revealed some 20 amplified chromosome regions in breast tumors. Positional cloning efforts are underway to identify the critical gene(s) in each amplified region. To date, genes known to be amplified frequently in breast cancers include *myc* (8q24), *ccnd1* (11q13), and *erbB2* (17q12-q21) (for review, see Bièche and Lidereau, 1995).

Amplification of the *myc*, *ccnd1*, and *erbB2* proto-oncogenes should have clinical relevance in breast cancer, since independent studies have shown that these alterations can be used to identify sub-populations with a worse prognosis (Berns *et al.*, 1992; Schuurin *et al.*, 1992; Slamon *et al.*, 1987). Muss *et al.* (1994) suggested that these gene alterations may also be useful for the prediction and assessment of the efficacy of adjuvant chemotherapy and hormone therapy.

However, published results diverge both in terms of the frequency of these alterations and their clinical value. For instance, over 500 studies in 10 years have failed to resolve the controversy

surrounding the link suggested by Slamon *et al.* (1987) between *erbB2* amplification and disease progression. These discrepancies are partly due to the clinical, histological and ethnic heterogeneity of breast cancer, but technical considerations are also probably involved.

Specific genes (DNA) were initially quantified in tumor cells by means of blotting procedures such as Southern and slot blotting. These batch techniques require large amounts of DNA (5–10 µg/reaction) to yield reliable quantitative results. Furthermore, meticulous care is required at all stages of the procedures to generate blots of sufficient quality for reliable dosage analysis. Recently, PCR has proven to be a powerful tool for quantitative DNA analysis, especially with minimal starting quantities of tumor samples (small, early-stage tumors and formalin-fixed, paraffin-embedded tissues).

Quantitative PCR can be performed by evaluating the amount of product either after a given number of cycles (end-point quantitative PCR) or after a varying number of cycles during the exponential phase (kinetic quantitative PCR). In the first case, an internal standard distinct from the target molecule is required to ascertain PCR efficiency. The method is relatively easy but implies generating, quantifying and storing an internal standard for each gene studied. Nevertheless, it is the most frequently applied method to date.

One of the major advantages of the kinetic method is its rapidity in quantifying a new gene, since no internal standard is required (an external standard curve is sufficient). Moreover, the kinetic method has a wide dynamic range (at least 5 orders of magnitude), giving an accurate value for samples differing in their copy number. Unfortunately, the method is cumbersome and has therefore been rarely used. It involves aliquot sampling of each assay mix at regular intervals and quantifying, for each aliquot, the amplification product. Interest in the kinetic method has been stimulated by a novel approach using fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real time (Gibson *et al.*, 1996; Heid *et al.*, 1996). The TaqMan reaction is based on the 5' nuclease assay first described by Holland *et al.* (1991). The latter uses the 5' nuclease activity of Taq polymerase to cleave a specific fluorogenic oligonucleotide probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee *et al.*, 1993). One fluorescent dye, co-valently linked to the 5' end of the oligonucleotide, serves as a reporter [FAM (*i.e.*, 6-carboxy-fluorescein)] and its emission spectrum is quenched by a second fluorescent dye, TAMRA (*i.e.*, 6-carboxy-tetramethyl-rhodamine) attached to the 3' end. During the extension phase of the PCR

Grant sponsors: Association Pour la Recherche sur le Cancer and Ministère de l'Enseignement Supérieur et de la Recherche.

*Correspondence to: Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, 4 Avenue de l'Observatoire, F-75006 Paris, France. Fax: (33)1-4407-1754. E-mail: mvidauid@teaser.fr

Received 2 May 1998; Revised 30 June 1998

cycle, the fluorescent hybridization probe is hydrolyzed by the 5'-3' nucleolytic activity of DNA polymerase. Nuclease degradation of the probe releases the quenching of FAM fluorescence emission, resulting in an increase in peak fluorescence emission. The fluorescence signal is normalized by dividing the emission intensity of the reporter dye (FAM) by the emission intensity of a reference dye (i.e., ROX, 6-carboxy-X-rhodamine) included in TaqMan buffer, to obtain a ratio defined as the R_n (normalized reporter) for a given reaction tube. The use of a sequence detector enables the fluorescence spectra of all 96 wells of the thermal cycler to be measured continuously during PCR amplification.

The real-time PCR method offers several advantages over other current quantitative PCR methods (Celi *et al.*, 1994): (i) the probe-based homogeneous assay provides a real-time method for detecting only specific amplification products, since specific hybridization of both the primers and the probe is necessary to generate a signal; (ii) the C_t (threshold cycle) value used for quantification is measured when PCR amplification is still in the log phase of PCR product accumulation. This is the main reason why C_t is a more reliable measure of the starting copy number than are end-point measurements, in which a slight difference in a limiting component can have a drastic effect on the amount of product; (iii) use of C_t values gives a wider dynamic range (at least 5 orders of magnitude), reducing the need for serial dilution; (iv) The real-time PCR method is run in a closed-tube system and requires no post-PCR sample handling, thus avoiding potential contamination; (v) the system is highly automated, since the instrument continuously measures fluorescence in all 96 wells of the thermal cycler during PCR amplification and the corresponding software processes, and analyzes the fluorescence data; (vi) the assay is rapid, as results are available just one minute after thermal cycling is complete; (vii) the sample throughput of the method is high, since 96 reactions can be analyzed in 2 hr.

Here, we applied this semi-automated procedure to determine the copy numbers of the 3 most frequently amplified genes in breast tumors (*myc*, *ccnd1* and *erbB2*), as well as 2 genes (*alb* and *app*) located in a chromosome region in which no genetic changes have been observed in breast tumors. The results for 108 breast tumors were compared with previous Southern-blot data for the same samples.

MATERIAL AND METHODS

Tumor and blood samples

Samples were obtained from 108 primary breast tumors removed surgically from patients at the Centre René Huguénin; none of the patients had undergone radiotherapy or chemotherapy. Immediately after surgery, the tumor samples were placed in liquid nitrogen until extraction of high-molecular-weight DNA. Patients were included in this study if the tumor sample used for DNA preparation contained more than 60% of tumor cells (histological analysis). A blood sample was also taken from 18 of the same patients.

DNA was extracted from tumor tissue and blood leukocytes according to standard methods.

Real-time PCR

Theoretical basis. Reactions are characterized by the point during cycling when amplification of the PCR product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the genomic DNA target, the earlier a significant increase in fluorescence is observed. The parameter C_t (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The target gene copy number in unknown samples is quantified by measuring C_t and by using a standard curve to determine the starting copy number. The precise amount of genomic DNA (based on optical density) and its quality (i.e., lack

of extensive degradation) are both difficult to assess. We therefore also quantified a control gene (*alb*) mapping to chromosome region 4q11-q13, in which no genetic alterations have been found in breast-tumor DNA by means of CGH (Kallioniemi *et al.*, 1994).

Thus, the ratio of the copy number of the target gene to the copy number of the *alb* gene normalizes the amount and quality of genomic DNA. The ratio defining the level of amplification, is termed "N", and is determined as follows:

$$N = \frac{\text{copy number of target gene (app, myc, ccnd1, erbB2)}}{\text{copy number of reference gene (alb)}}$$

Primers, probes, reference human genomic DNA and PCR consumables. Primers and probes were chosen with the assistance of the computer programs Oligo 4.0 (National Biosciences, Plymouth, MN), EuGene (Daniben Systems, Cincinnati, OH) and Primer Express (Perkin-Elmer Applied Biosystems, Foster City, CA).

Primers were purchased from DNAScience (Malvern, PA) and probes from Perkin-Elmer Applied Biosystems.

Nucleotide sequences for the oligonucleotide hybridization probes and primers are available on request.

The TaqMan PCR Core reagent kit, MicroAmp optical tubes, and MicroAmp caps were from Perkin-Elmer Applied Biosystems.

Standard-curve construction. The kinetic method requires a standard curve. The latter was constructed with serial dilutions of specific PCR products, according to Piatak *et al.* (1993). In practice, each specific PCR product was obtained by amplifying 20 ng of a standard human genomic DNA (Boehringer, Mannheim, Germany) with the same primer pairs as those used later for real-time quantitative PCR. The 5 PCR products were purified using MicroSpin S-400 HR columns (Pharmacia, Uppsala, Sweden) electrophoresed through an acrylamide gel and stained with ethidium bromide to check their quality. The PCR products were then quantified spectrophotometrically and pooled, and serially diluted 10-fold in mouse genomic DNA (Clontech, Palo Alto, CA) at a constant concentration of 2 ng/ μ l. The standard curve used for real-time quantitative PCR was based on serial dilutions of the pool of PCR products ranging from 10^{-7} (10^3 copies of each gene) to 10^{-10} (10^0 copies). This series of diluted PCR products was aliquoted and stored at -80°C until use.

The standard curve was validated by analyzing 2 known quantities of calibrator human genomic DNA (20 ng and 50 ng).

PCR amplification. Amplification mixes (50 μ l) contained the sample DNA (around 20 ng, around 6600 copies of disomic genes), $10\times$ TaqMan buffer (5 μ l), 200 μ M dATP, dCTP, dGTP, and 400 μ M dUTP, 5 mM MgCl_2 , 1.25 units of AmpliTaq Gold, 0.5 units of AmpErase uracil N-glycosylase (UNG), 200 nM each primer and 100 nM probe. The thermal cycling conditions comprised 2 min at 50°C and 10 min at 95°C . Thermal cycling consisted of 40 cycles at 95°C for 15 s and 65°C for 1 min. Each assay included: a standard curve (from 10^3 to 10^0 copies) in duplicate, a no-template control, 20 ng and 50 ng of calibrator human genomic DNA (Boehringer) in triplicate, and about 20 ng of unknown genomic DNA in triplicate (26 samples can thus be analyzed on a 96-well microplate). All samples with a coefficient of variation (CV) higher than 10% were retested.

All reactions were performed in the ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems), which detects the signal from the fluorogenic probe during PCR.

Equipment for real-time detection. The 7700 system has a built-in thermal cycler and a laser directed via fiber optical cables to each of the 96 sample wells. A charge-coupled-device (CDD) camera collects the emission from each sample and the data are analyzed automatically. The software accompanying the 7700 system calculates C_t and determines the starting copy number in the samples.

Determination of gene amplification. Gene amplification was calculated as described above. Only samples with an N value higher than 2 were considered to be amplified.

RESULTS

To validate the method, real-time PCR was performed on genomic DNA extracted from 108 primary breast tumors, and 18 normal leukocyte DNA samples from some of the same patients. The target genes were the *myc*, *ccnd1* and *erbB2* proto-oncogenes, and the β -amyloid precursor protein gene (*app*), which maps to a chromosome region (21q21.2) in which no genetic alterations have been found in breast tumors (Kallioniemi *et al.*, 1994). The reference disomic gene was the albumin gene (*alb*, chromosome 4q11-q13).

Validation of the standard curve and dynamic range of real-time PCR

The standard curve was constructed from PCR products serially diluted in genomic mouse DNA at a constant concentration of 2 ng/ μ l. It should be noted that the 5 primer pairs chosen to analyze the 5 target genes do not amplify genomic mouse DNA (data not shown). Figure 1 shows the real-time PCR standard curve for the *alb* gene. The dynamic range was wide (at least 4 orders of magnitude), with samples containing as few as 10^2 copies or as many as 10^5 copies.

Copy-number ratio of the 2 reference genes (*app* and *alb*)

The *app* to *alb* copy-number ratio was determined in 18 normal leukocyte DNA samples and all 108 primary breast-tumor DNA

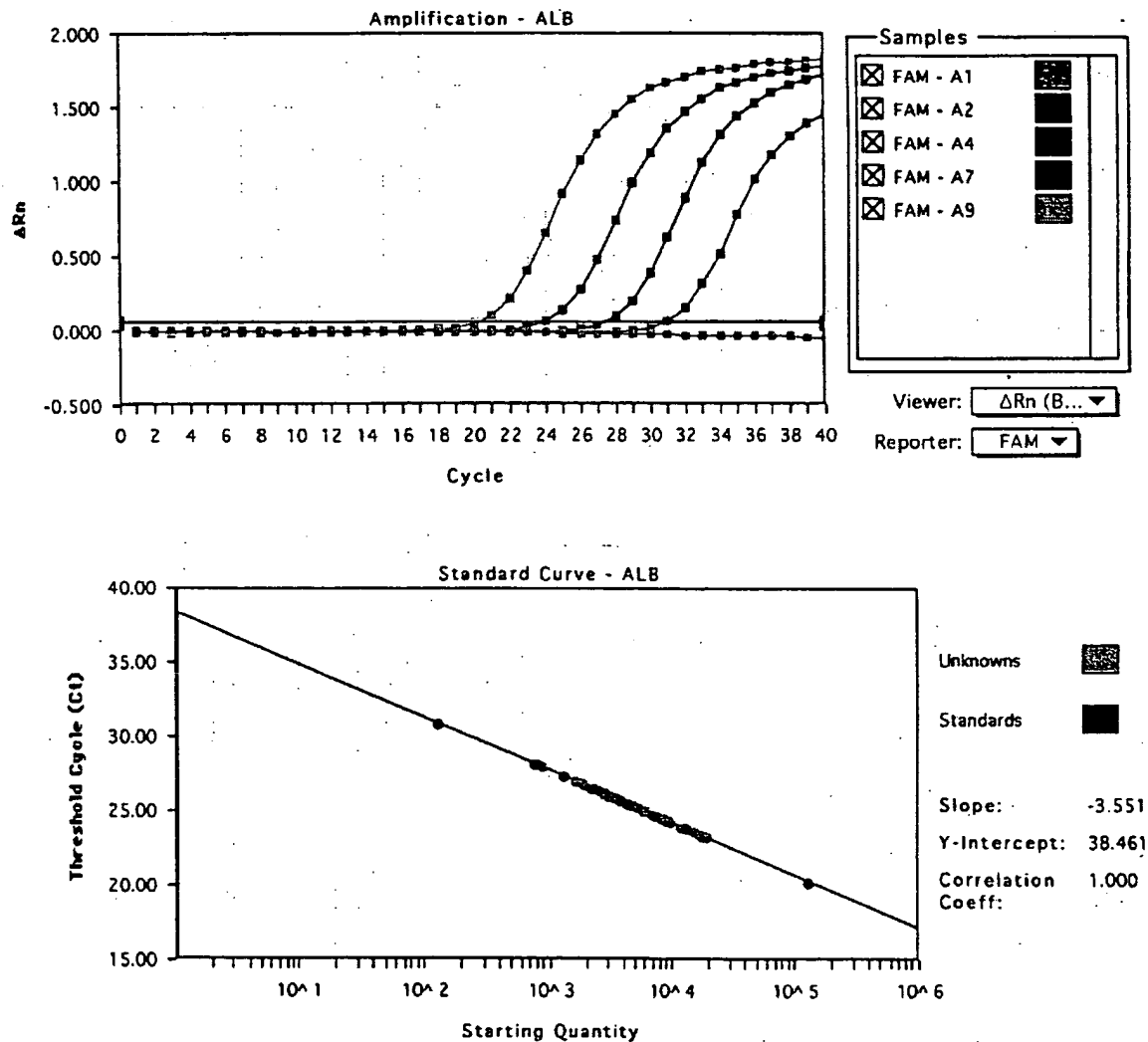


FIGURE 1 – Albumin (*alb*) gene dosage by real-time PCR. Top: Amplification plots for reactions with starting *alb* gene copy number ranging from 10^5 (A9), 10^4 (A7), 10^3 (A4) to 10^2 (A2) and a no-template control (A1). Cycle number is plotted vs. change in normalized reporter signal (ΔRn). For each reaction tube, the fluorescence signal of the reporter dye (FAM) is divided by the fluorescence signal of the passive reference dye (ROX), to obtain a ratio defined as the normalized reporter signal (Rn). ΔRn represents the normalized reporter signal (Rn) minus the baseline signal established in the first 15 PCR cycles. ΔRn increases during PCR as *alb* PCR product copy number increases until the reaction reaches a plateau. C_t (threshold cycle) represents the fractional cycle number at which a significant increase in Rn above a baseline signal (horizontal black line) can first be detected. Two replicate plots were performed for each standard sample, but the data for only one are shown here. Bottom: Standard curve plotting log starting copy number vs. C_t (threshold cycle). The black dots represent the data for standard samples plotted in duplicate and the red dots the data for unknown genomic DNA samples plotted in triplicate. The standard curve shows 4 orders of linear dynamic range.

samples. We selected these 2 genes because they are located in 2 chromosome regions (*app*, 21q21.2; *alb*, 4q11-q13) in which no obvious genetic changes (including gains or losses) have been observed in breast cancers (Kallioniemi *et al.*, 1994). The ratio for the 18 normal leukocyte DNA samples fell between 0.7 and 1.3 (mean 1.02 ± 0.21), and was similar for the 108 primary breast-tumor DNA samples (0.6 to 1.6 , mean 1.06 ± 0.25), confirming that *alb* and *app* are appropriate reference disomic genes for breast-tumor DNA. The low range of the ratios also confirmed that the nucleotide sequences chosen for the primers and probes were not polymorphic, as mismatches of their primers or probes with the subject's DNA would have resulted in differential amplification.

myc, *ccnd1* and *erbB2* gene dose in normal leukocyte DNA

To determine the cut-off point for gene amplification in breast-cancer tissue, 18 normal leukocyte DNA samples were tested for the gene dose (N), calculated as described in "Material and Methods". The N value of these samples ranged from 0.5 to 1.3 (mean 0.84 ± 0.22) for *myc*, 0.7 to 1.6 (mean 1.06 ± 0.23) for *ccnd1* and 0.6 to 1.3 (mean 0.91 ± 0.19) for *erbB2*. Since N values for *myc*, *ccnd1* and *erbB2* in normal leukocyte DNA consistently fell between 0.5 and 1.6, values of 2 or more were considered to represent gene amplification in tumor DNA.

myc, *ccnd1* and *erbB2* gene dose in breast-tumor DNA

myc, *ccnd1* and *erbB2* gene copy numbers in the 108 primary breast tumors are reported in Table I. Extra copies of *ccnd1* were more frequent (23%, 25/108) than extra copies of *erbB2* (15%, 16/108) and *myc* (10%, 11/108), and ranged from 2 to 18.6 for *ccnd1*, 2 to 15.1 for *erbB2*, and only 2 to 4.6 for the *myc* gene. Figure 2 and Table II represent tumors in which the *ccnd1* gene was amplified 16-fold (T145), 6-fold (T133) and non-amplified (T118). The 3 genes were never found to be co-amplified in the same tumor. *erbB2* and *ccnd1* were co-amplified in only 3 cases, *myc* and *ccnd1* in 2 cases and *myc* and *erbB2* in 1 case. This favors the hypothesis that gene amplifications are independent events in breast cancer. Interestingly, 5 tumors showed a decrease of at least 50% in the *erbB2* copy number ($N < 0.5$), suggesting that they bore deletions of the 17q21 region (the site of *erbB2*). No such decrease in copy number was observed with the other 2 proto-oncogenes.

Comparison of gene dose determined by real-time quantitative PCR and Southern-blot analysis

Southern-blot analysis of *myc*, *ccnd1* and *erbB2* amplifications had previously been done on the same 108 primary breast tumors. A perfect correlation between the results of real-time PCR and Southern blot was obtained for tumors with high copy numbers ($N \geq 5$). However, there were cases (1 *myc*, 6 *ccnd1* and 4 *erbB2*) in which real-time PCR showed gene amplification whereas Southern-blot did not, but these were mainly cases with low extra copy numbers (N from 2 to 2.9).

DISCUSSION

The clinical applications of gene amplification assays are currently limited, but would certainly increase if a simple, standardized and rapid method were perfected. Gene amplification status has been studied mainly by means of Southern blotting, but this method is not sensitive enough to detect low-level gene amplification nor accurate enough to quantify the full range of amplification values. Southern blotting is also time-consuming, uses radioactive

reagents and requires relatively large amounts of high-quality genomic DNA, which means it cannot be used routinely in many laboratories. An amplification step is therefore required to determine the copy number of a given target gene from minimal quantities of tumor DNA (small early-stage tumors, cytopuncture specimens or formalin-fixed, paraffin-embedded tissues).

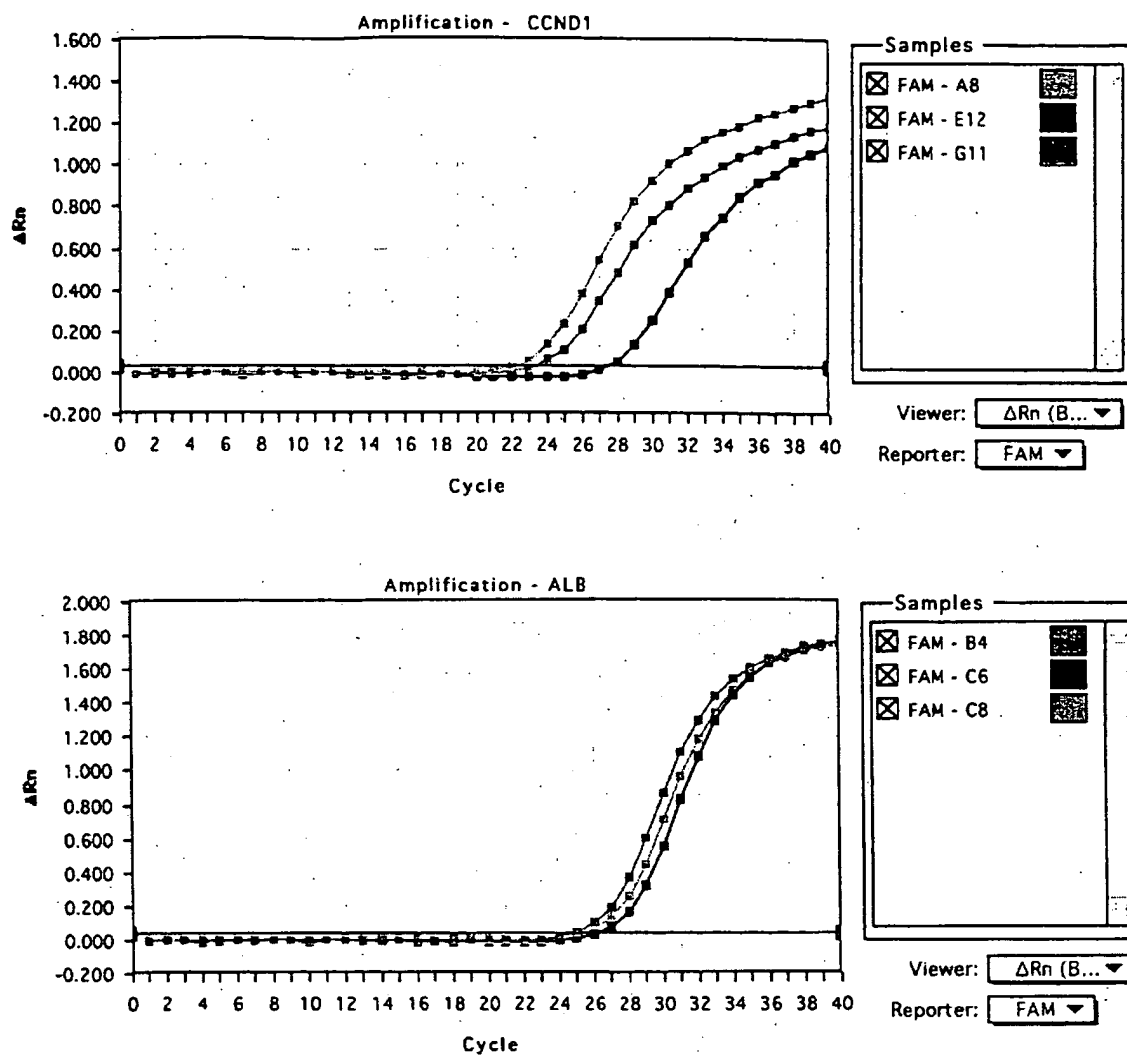
In this study, we validated a PCR method developed for the quantification of gene over-representation in tumors. The method, based on real-time analysis of PCR amplification, has several advantages over other PCR-based quantitative assays such as competitive quantitative PCR (Celi *et al.*, 1994). First, the real-time PCR method is performed in a closed-tube system, avoiding the risk of contamination by amplified products. Re-amplification of carryover PCR products in subsequent experiments can also be prevented by using the enzyme uracil N-glycosylase (UNG) (Longo *et al.*, 1990). The second advantage is the simplicity and rapidity of sample analysis, since no post-PCR manipulations are required. Our results show that the automated method is reliable. We found it possible to determine, in triplicate, the number of copies of a target gene in more than 100 tumors per day. Third, the system has a linear dynamic range of at least 4 orders of magnitude, meaning that samples do not have to contain equal starting amounts of DNA. This technique should therefore be suitable for analyzing formalin-fixed, paraffin-embedded tissues. Fourth, and above all, real-time PCR makes DNA quantification much more precise and reproducible, since it is based on C_t values rather than end-point measurement of the amount of accumulated PCR product. Indeed, the ABI Prism 7700 Sequence Detection System enables C_t to be calculated when PCR amplification is still in the exponential phase and when none of the reaction components is rate-limiting. The within-run CV of the C_t value for calibrator human DNA (5 replicates) was always below 5%, and the between-assay precision in 5 different runs was always below 10% (data not shown). In addition, the use of a standard curve is not absolutely necessary, since the copy number can be determined simply by comparing the C_t ratio of the target gene with that of reference genes. The results obtained by the 2 methods (with and without a standard curve) are similar in our experiments (data not shown). Moreover, unlike competitive quantitative PCR, real-time PCR does not require an internal control (the design and storage of internal controls and the validation of their amplification efficiency is laborious).

The only potential disadvantage of real-time PCR, like all other PCR-based methods and solid-matrix blotting techniques (Southern blots and dot blots) is that it cannot avoid dilution artifacts inherent in the extraction of DNA from tumor cells contained in heterogeneous tissue specimens. Only FISH and immunohistochemistry can measure alterations on a cell-by-cell basis (Pauletti *et al.*, 1996; Slamon *et al.*, 1989). However, FISH requires expensive equipment and trained personnel and is also time-consuming. Moreover, FISH does not assess gene expression and therefore cannot detect cases in which the gene product is over-expressed in the absence of gene amplification, which will be possible in the future by real-time quantitative RT-PCR. Immunohistochemistry is subject to considerable variations in the hands of different teams, owing to alterations of target proteins during the procedure, the different primary antibodies and fixation methods used and the criteria used to define positive staining.

The results of this study are in agreement with those reported in the literature. (i) Chromosome regions 4q11-q13 and 21q21.2 (which bear *alb* and *app*, respectively) showed no genetic alterations in the breast-cancer samples studied here, in keeping with the results of CGH (Kallioniemi *et al.*, 1994). (ii) We found that amplifications of these 3 oncogenes were independent events, as reported by other teams (Berns *et al.*, 1992; Borg *et al.*, 1992). (iii) The frequency and degree of *myc* amplification in our breast tumor DNA series were lower than those of *ccnd1* and *erbB2* amplification, confirming the findings of Borg *et al.* (1992) and Courjal *et al.* (1997). (iv) The maxima of *ccnd1* and *erbB2* over-representation were 18-fold and 15-fold, also in keeping with earlier results (about

TABLE I - DISTRIBUTION OF AMPLIFICATION LEVEL (N) FOR *myc*, *ccnd1* AND *erbB2* GENES IN 108 HUMAN BREAST TUMORS

Gene	Amplification level (N)			
	<0.5	0.5-1.9	2-4.9	≥ 5
<i>myc</i>	0	97 (89.8%)	11 (10.2%)	0
<i>ccnd1</i>	0	83 (76.9%)	17 (15.7%)	8 (7.4%)
<i>erbB2</i>	5 (4.6%)	87 (80.6%)	8 (7.4%)	8 (7.4%)



Tumor	CCND1		ALB	
	C_t	Copy number	C_t	Copy number
■ T118	27.3	4605	26.5	4365
▨ T133	23.2	61659	25.2	10092
▩ T145	22.1	125892	25.6	7762

FIGURE 2 – *ccnd1* and *alb* gene dosage by real-time PCR in 3 breast tumor samples: T118 (E12, C6, black squares), T133 (G11, B4, red squares) and T145 (A8, C8, blue squares). Given the C_t of each sample, the initial copy number is inferred from the standard curve obtained during the same experiment. Triplicate plots were performed for each tumor sample, but the data for only one are shown here. The results are shown in Table II.

30-fold maximum) (Berns *et al.*, 1992; Borg *et al.*, 1992; Courjal *et al.*, 1997). (v) The *erbB2* copy numbers obtained with real-time PCR were in good agreement with data obtained with other quantitative PCR-based assays in terms of the frequency and degree of amplification (An *et al.*, 1995; Deng *et al.*, 1996; Valeron

et al., 1996). Our results also correlate well with those recently published by Gelmini *et al.* (1997), who used the TaqMan system to measure *erbB2* amplification in a small series of breast tumors ($n = 25$), but with an instrument (LS-50B luminescence spectrometer, Perkin-Elmer Applied Biosystems) which only allows end-

TABLE II - EXAMPLES OF *ccnd1* GENE DOSAGE RESULTS FROM 3 BREAST TUMORS¹

Tumor	<i>ccnd1</i>			<i>alb</i>			<i>Nccnd1/alb</i>
	Copy number	Mean	SD	Copy number	Mean	SD	
T118	4525			4223			
	4605	4603	77	4365	4325	89	1.06
	4678			4387			
T133	59821			9787			
	61659	61100	1111	10092	10137	375	6.03
	61821			10533			
T145	128563			7321			
	125892	125392	3448	7762	7672	316	16.34
	121722			7933			

¹For each sample, 3 replicate experiments were performed and the mean and the standard deviation (SD) was determined. The level of *ccnd1* gene amplification (*Nccnd1/alb*) is determined by dividing the average *ccnd1* copy number value by the average *alb* copy number value.

point measurement of fluorescence intensity. Here we report *myc* and *ccnd1* gene dosage in breast cancer by means of quantitative PCR. (vi) We found a high degree of concordance between real-time quantitative PCR and Southern blot analysis in terms of gene amplification, especially for samples with high copy numbers (≥ 5 -fold). The slightly higher frequency of gene amplification (especially *ccnd1* and *erbB2*) observed by means of real-time quantitative PCR as compared with Southern-blot analysis may be explained by the higher sensitivity of the former method. However, we cannot rule out the possibility that some tumors with a few extra

gene copies observed in real-time PCR had additional copies of an arm or a whole chromosome (trisomy, tetrasomy or polysomy) rather than true gene amplification. These 2 types of genetic alteration (polysomy and gene amplification) could be easily distinguished in the future by using an additional probe located on the same chromosome arm, but some distance from the target gene. It is noteworthy that high gene copy numbers have the greatest prognostic significance in breast carcinoma (Borg *et al.*, 1992; Slamon *et al.*, 1987).

Finally, this technique can be applied to the detection of gene deletion as well as gene amplification. Indeed, we found a decreased copy number of *erbB2* (but not of the other 2 proto-oncogenes) in several tumors; *erbB2* is located in a chromosome region (17q21) reported to contain both deletions and amplifications in breast cancer (Bièche and Lidereau, 1995).

In conclusion, gene amplification in various cancers can be used as a marker of pre-neoplasia, also for early diagnosis of cancer, staging, prognostication and choice of treatment. Southern blotting is not sufficiently sensitive, and FISH is lengthy and complex. Real-time quantitative PCR overcomes both these limitations, and is a sensitive and accurate method of analyzing large numbers of samples in a short time. It should find a place in routine clinical gene dosage.

ACKNOWLEDGEMENTS

RL is a research director at the Institut National de la Santé et de la Recherche Médicale (INSERM). We thank the staff of the Centre René Huguenin for assistance in specimen collection and patient care.

REFERENCES

- AN, H.X., NIEDERACHER, D., BECKMANN, M.W., GÖHRING, U.J., SCHARL, A., PICARD, F., VAN ROEYEN, C., SCHNÜRCH, H.G. and BENDER, H.G., *erbB2* gene amplification detected by fluorescent differential polymerase chain reaction in paraffin-embedded breast carcinoma tissues. *Int. J. Cancer (Pred. Oncol.)*, 64, 291-297 (1995).
- BERNS, E.M.J.J., KLIN, J.G.M., VAN PUTTEN, W.L.J., VAN STAVEREN, I.L., PORTENG, H. and FOEKENS, J.A., *c-myc* amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res.*, 52, 1107-1113 (1992).
- BIÈCHE, I. and LIDEREAU, R., Genetic alterations in breast cancer. *Genes Chrom. Cancer*, 14, 227-251 (1995).
- BORG, A., BALDETORP, B., FERNO, M., OLSSON, H. and SIGURDSSON, H., *c-myc* amplification is an independent prognostic factor in post-menopausal breast cancer. *Int. J. Cancer*, 51, 687-691 (1992).
- CELI, F.S., COHEN, M.M., ANTONARAKIS, S.E., WERTHEIMER, E., ROTH, J. and SHULDINER, A.R., Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (qd-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics*, 21, 304-310 (1994).
- COURJAL, F., CUNY, M., SIMONY-LAFONTAINE, J., LOUASSON, G., SPEISER, P., ZEILLINGER, R., RODRIGUEZ, C. and THEILLET, C., Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.*, 57, 4360-4367 (1997).
- DENG, G., YU, M., CHEN, L.C., MOORE, D., KURISU, W., KALLIONIEMI, A., WALDMAN, F.M., COLLINS, C. and SMITH, H.S., Amplifications of oncogene *erbB-2* and chromosome 20q in breast cancer determined by differentially competitive polymerase chain reaction. *Breast Cancer Res. Treat.*, 40, 271-281 (1996).
- GELMINI, S., ORIANDO, C., SESTINI, R., VONA, G., PINZANI, P., RUOCO, L. and PAZZAGLI, M., Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure *c-erbB-2* oncogene amplification. *Clin. Chem.*, 43, 752-758 (1997).
- GIBSON, U.E.M., HEID, C.A. and WILLIAMS, P.M., A novel method for real-time quantitative RT-PCR. *Genome Res.*, 6, 995-1001 (1996).
- HEID, C.A., STEVENS, J., LIVAK, K.J. and WILLIAMS, P.M., Real-time quantitative PCR. *Genome Res.*, 6, 986-994 (1996).
- HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. and GELFAND, D.H., Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. nat. Acad. Sci. (Wash.)*, 88, 7276-7280 (1991).
- KALLIONIEMI, A., KALLIONIEMI, O.P., PIPER, J., TANNER, M., STOKKES, T., CHEN, L., SMITH, H.S., PINKEL, D., GRAY, J.W. and WALDMAN, F.M., Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. nat. Acad. Sci. (Wash.)*, 91, 2156-2160 (1994).
- LEE, L.G., CONNELL, C.R. and BIOCH, W., Allelic discrimination by nick-translation PCR with fluorogenic probe. *Nucleic Acids Res.*, 21, 3761-3766 (1993).
- LONGO, N., BERNINGER, N.S. and HARTLEY, J.L., Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene*, 93, 125-128 (1990).
- MUSS, H.B., THOR, A.D., BERRY, D.A., KUTE, T., LIU, E.T., KOERNER, F., CIRINCIONE, C.T., BUDMAN, D.R., WOOD, W.C., BARCOS, M. and HENDERSON, I.C., *c-erbB-2* expression and response to adjuvant therapy in women with node-positive early breast cancer. *New Engl. J. Med.*, 330, 1260-1266 (1994).
- PAULETTI, G., GODOLPHIN, W., PRESS, M.F. and SALMON, D.J., Detection and quantification of *HER-2/neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene*, 13, 63-72 (1996).
- PIATAK, M., LUK, K.C., WILLIAMS, B. and LIFSON, J.D., Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Biotechniques*, 14, 70-80 (1993).
- SCHUURING, E., VERHOEVEN, E., VAN TINTEREN, H., PETERSE, J.L., NUNNIK, B., THUNNISSEN, F.B.J.M., DEVILLE, P., CORNELISSE, C.J., VAN DE VIVER, M.J., MOOI, W.J. and MICHALIDES, R.J.A.M., Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.*, 52, 5229-5234 (1992).
- SLAMON, D.J., CLARK, G.M., WONG, S.G., LEVIN, W.S., ULLRICH, A. and MCGUIRE, W.L., Human breast cancer: correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science*, 235, 177-182 (1987).
- SLAMON, D.J., GODOLPHIN, W., JONES, L.A., HOLT, J.A., WONG, S.G., KEITH, D.E., LEVIN, W.J., STUART, S.G., UDOWE, J., ULLRICH, A. and PRESS, M.F., Studies of the *HER-2/neu* proto-oncogene in human breast and ovarian cancer. *Science*, 244, 707-712 (1989).
- VALERON, P.F., CHIRINO, R., FERNANDEZ, L., TORRES, S., NAVARRO, D., AGUIAR, J., CABRERA, J.J., DIAZ-CHICO, B.N. and DIAZ-CHICO, J.C., Validation of a differential PCR and an ELISA procedure in studying *HER-2/neu* status in breast cancer. *Int. J. Cancer*, 65, 129-133 (1996).

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ashkenazi et al.
App. No. : 09/903,925
Filed : July 11, 2001
For : SECRETED AND
TRANSMEMBRANE
POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
Examiner : Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

(Date)

Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.
2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan[®] PCR or the fluorescence *in situ* hybridization (FISH) assays - is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi
Avi Ashkenazi, Ph.D.

Date: 9/15/03

CURRICULUM VITAE

Avi Ashkenazi

July 2003

Personal:

Date of birth: 29 November, 1956
Address: 1456 Tarrytown Street, San Mateo, CA 94402
Phone: (650) 578-9199 (home); (650) 225-1853 (office)
Fax: (650) 225-6443 (office)
Email: aa@gene.com

Education:

1983: B.S. in Biochemistry, with honors, Hebrew University, Israel
1986: Ph.D. in Biochemistry, Hebrew University, Israel

Employment:

1983-1986: Teaching assistant, undergraduate level course in Biochemistry
1985-1986: Teaching assistant, graduate level course on Signal Transduction
1986 - 1988: Postdoctoral fellow, Hormone Research Dept., UCSF, and
Developmental Biology Dept., Genentech, Inc., with J. Ramachandran
1988 - 1989: Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc.,
with D. Capon
1989 - 1993: Scientist, Molecular Biology Dept., Genentech, Inc.
1994 -1996: Senior Scientist, Molecular Oncology Dept., Genentech, Inc.
1996-1997: Senior Scientist and Interim director, Molecular Oncology Dept.,
Genentech, Inc.
1997-1990: Senior Scientist and preclinical project team leader, Genentech, Inc.
1999 -2002: Staff Scientist in Molecular Oncology, Genentech, Inc.
2002-present: Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

Awards:

1988: First prize, The Boehringer Ingelheim Award

Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* **34**, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* **118**, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* **50**, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* **14**, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* **210**, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotrophic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* **121**, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* **236**, 600-605 (1987).
8. Peralta, E., Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* **6**, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* **238**, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* **37**, 119-129 (1988).
11. Peralta, E., Ashkenazi, A., Winslow, J., Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* **334**, 434-437 (1988).
12. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* **56**, 487-493 (1989).
13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* **340**, 146-150 (1989).
14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* **245**, 743-745 (1989).
15. Ashkenazi, A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodeficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* **87**, 7150-7154 (1990).
16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* **29**, 9885-9891 (1990).
17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-gp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* **88**, 7056-7060 (1991).
18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari, I., Pennica, D., Goeddel, D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* **88**, 10535-10539 (1991).
19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* **66**, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel, R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine_{1B} receptor. *J. Biol. Chem.* **267**, 5735-5738 (1992).
21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* **267**, 5747-5750 (1992).
22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).
23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT_{1B} receptors. *Nature* **360**, 161-163 (1992).
24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon γ by an interferon γ receptor immunoadhesin. *Immunology* **79**, 594-599 (1993).
25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, A., Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* **32**, 3131-3138. (1993).
26. Barford, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* **268**, 26059-26062 (1993).
27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3⁺ effectors to kill HIV-1-infected cells. *J. Immunol.* **153**, 4268-4280 (1994).
28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by interferon γ and correction by interferon γ receptor immunoadhesin. *Blood* **83**, 911-915 (1994).
29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., and Ashkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* **152**, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkylation. *Bioconj. Chem.* **5**, 133-140 (1994).
31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* **170**, 1323-1326 (1994).
32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* **31**, 1335-1344 (1994).
33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* **31**, 1345-1351 (1994).
34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT₂ receptor specifies effector coupling. *J. Neurochem.* **64**, 1440-1447 (1995).
35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN- γ receptor β chain expression in T helper cell subsets. *Science* **270**, 1215-1218 (1995).
36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* **95**, 1947-1952 (1995).
37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon γ signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA*. **92**, 5401-5405 (1995).
38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal *E. coli* bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* **156**, 2221-2230 (1996).
39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 1669-1676 (1996).
41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF- κ B and AP-1 transcription factors. *J. Biol. Chem.* 272, 14029-14032 (1997).
42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* 277, 818-821 (1997).
43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* 7, 1003-1006 (1997).
44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* 427, 124-128 (1998).
46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* 160, 4098-4103 (1998).
47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* 396, 699-703 (1998).
48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* 162, 5616-5623 (1999).
49. Gurney, A. Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* 9, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155-162 (1999).
51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN- α receptor 2. *J. Immunol.* **163**, 766-773 (1999).
52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* **11**, 664-672 (1999).
53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* **265**, 479-483 (1999).
54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* **4**, 563–571 (1999).
55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633-640 (2000).
56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* **60**, 2611-2615 (2000).
57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611-620 (2000).
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., *Ashkenazi, A., and *Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* 10, 785-788 (2000).
60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* 29, 506-512 (2000).
61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D., Mounho, B., Hillan, K., Totpal, K., DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* 7, 383-385 (2001).
62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Marsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* 166, 4891-4898 (2001).
63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* 7, 1362-1369 (2001).
64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., *Ashkenazi, A., and *Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* 2, 632-637 (2001).
65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).
66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A. Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).
67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault, Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

Review articles:

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliège, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzymology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook (www.apnet.com/cytokinereference/).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. *Cytokine and Growth Factor Reviews* 14, 337-348 (2003).

Book:

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

Talks:

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Half-life Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoadhesins for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoadhesins: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficacy of TNF receptor immunoadhesin vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon- γ signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Francisco, CA, July 1995.
8. Immunoadhesins: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: a novel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philadelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philadelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14th Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Weizmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Switzerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

Issued Patents:

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6,046,048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft†§, Thomas Thykjaer¶, Frederic M. Waldman||, Hans Wolf**, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. *Molecular & Cellular Proteomics* 1:37–45, 2002.

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this

phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, cyclin d1, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)¹ have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q–, 11p–, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

From the †Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and **Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ‡AROS Applied Biotechnology ApS, Gustav Wiedsvvej 10, DK-8000 Aarhus C, Denmark, ||UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ‡‡Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

¹ The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

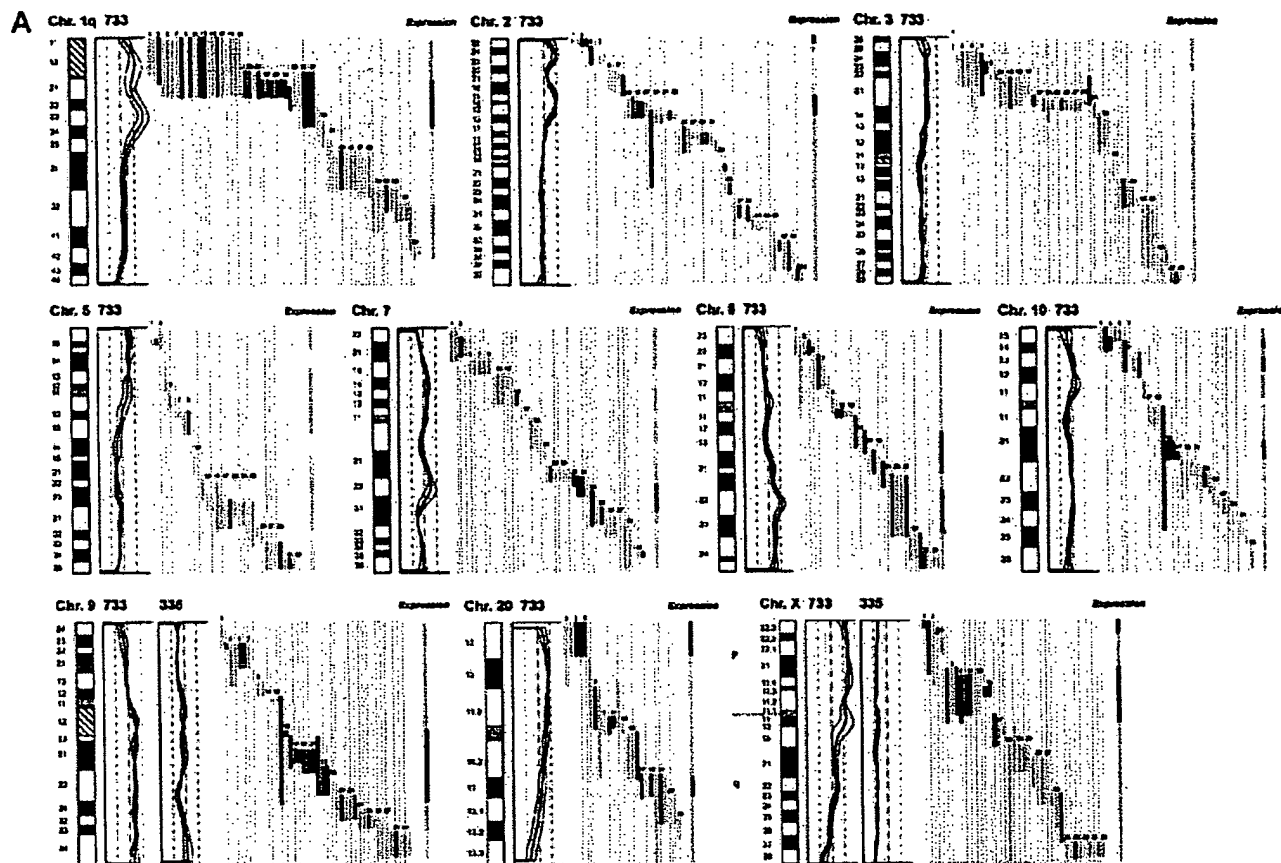


FIG. 1. DNA copy number and mRNA expression level. Shown from left to right are chromosome (Chr.), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. **A**, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. **B**, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (left). The bold curve in the ratio profile represents a mean of four chromosomes and is surrounded by thin curves indicating one standard deviation. The central vertical line (broken) indicates a ratio value of 1 (no change), and the vertical lines next to it (dotted) indicate a ratio of 0.5 (left) and 2.0 (right). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the right of the invasive tumor profile. The colored bars represent one gene each, identified by the running numbers above the bars (the name of the gene can be seen at www.MDL.DK/sdata.html). The bars indicate the purported location of the gene, and the colors indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (black), >2-fold decrease (blue), no significant change (orange). The bar to the far right, entitled *Expression* shows the resulting change in expression along the chromosome; the colors indicate that at least half of the genes were up-regulated (black), at least half of the genes down-regulated (blue), or more than half of the genes are unchanged (orange). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at -80°C . Total RNA was isolated using the RNeasy B RNA isolation method (WAK-Chemie Medical GMBH). poly(A)⁺ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript[®] choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the ME-GAscrip[®] *in vitro* transcription kit (Ambion). Biotin-labeled CTP and

UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94°C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6 \times SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95°C for 5 min, subsequently cooled to 40°C , and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40°C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6 \times SSPE-T at 25°C followed by 4 washes in 0.5 \times SSPE-T at 50°C . The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 $\mu\text{g}/\text{ml}$ (Molecular Probes) in 6 \times SSPE-T

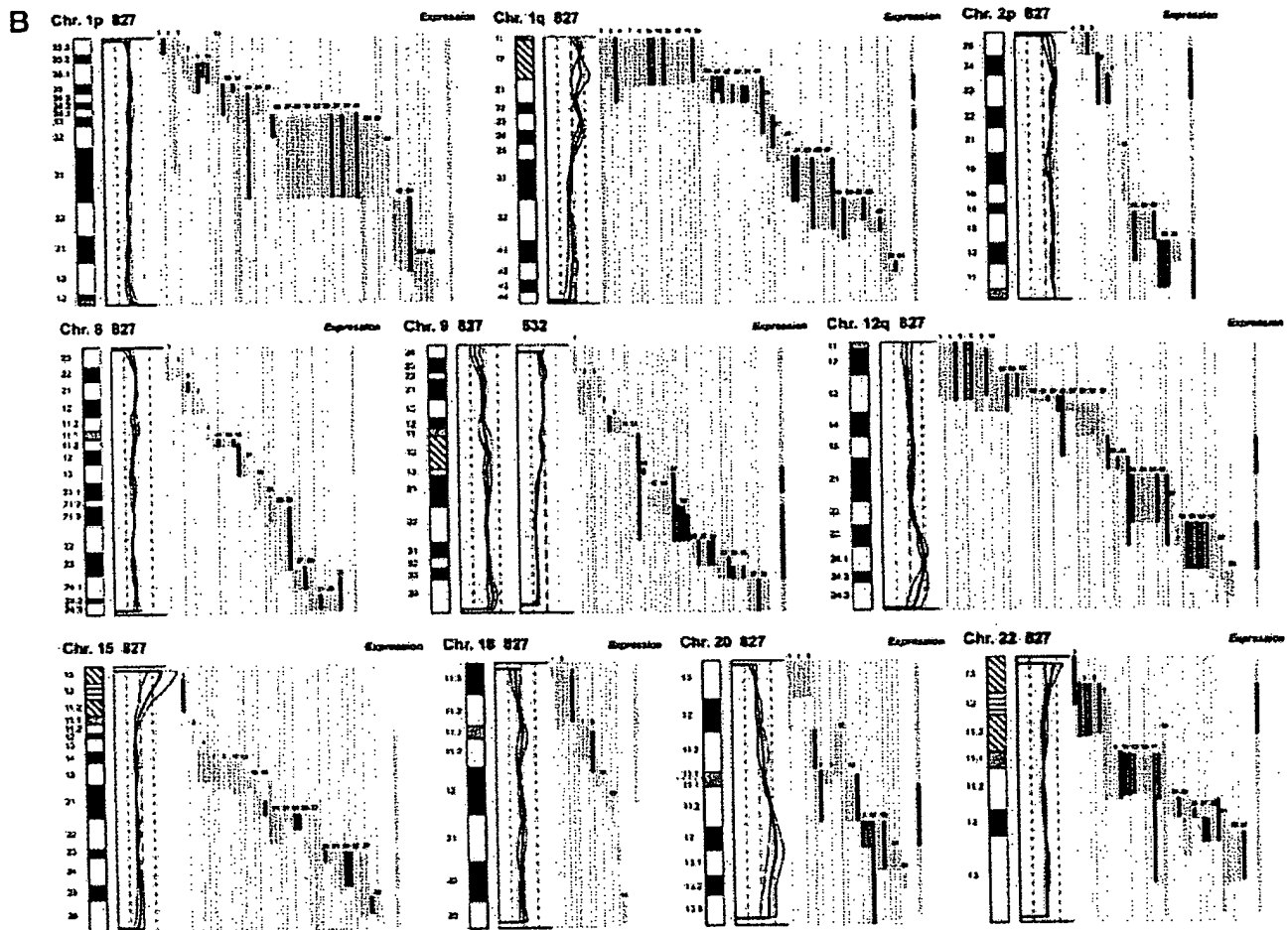


Fig. 1—continued

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μ l for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at -20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgl-bin/celis.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-

labeled reference DNA (200 ng), and human Cot-1 DNA (20 μ g) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μ g/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

Gene Copy Numbers, Transcripts, and Protein Levels

TABLE I
Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335 Expression change clusters	Concordance	CGH alterations	Tumor 827 vs. 532 Expression change clusters	Concordance
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down regulation 7 No change	17%
Expression change clusters	Tumor 733 vs. 335 CGH alterations	Concordance	Expression change clusters	Tumor 827 vs. 532 CGH alterations	Concordance
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22–q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22–24+, 2q14.1–qter–, 3q12–q13.3–, 6q12–q22–, 9q34+, 11q12–q13+, 17+, and 20q11.2–q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1A) and 20q12 in TCC 827 (Fig. 1B).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive versus the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger

than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21–q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, top). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, top). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, bottom). Furthermore, as a control we looked at areas with no alter-

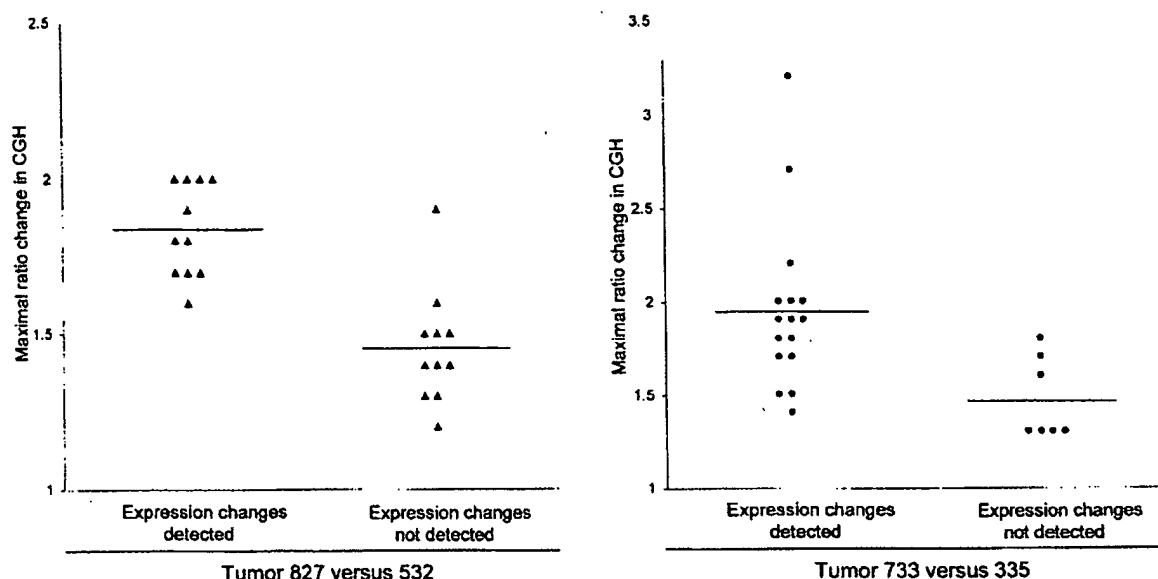


Fig. 2. Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring. The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (◆) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the *right* in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or

because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1A). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

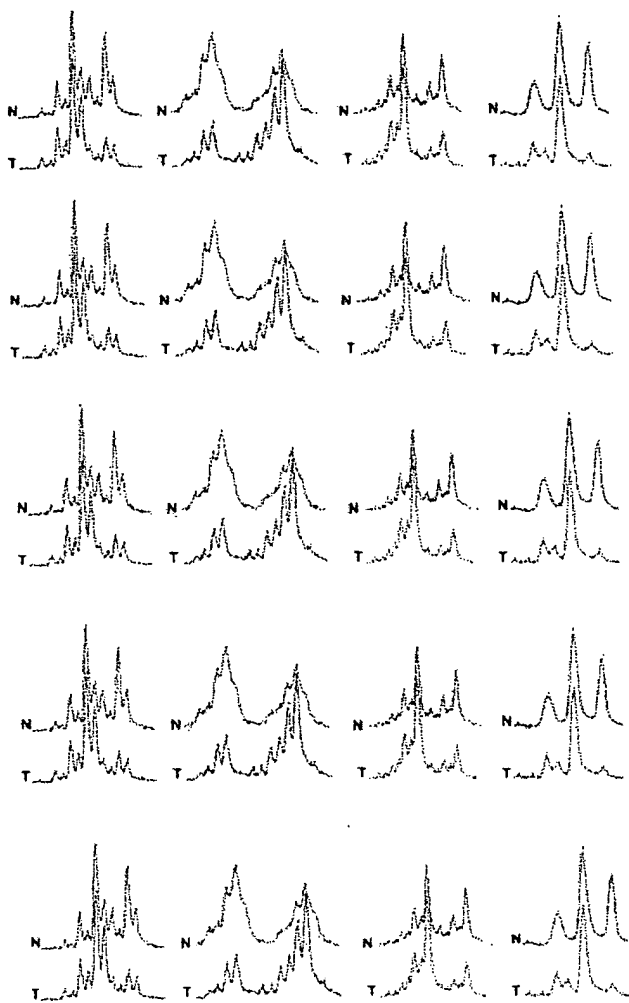


FIG. 3. Microsatellite analysis of loss of heterozygosity. Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (a) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (b) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (c) at chromosome 2p23 by D2S2251 close to general β -spectrin (gene number 11 on Fig. 1) and of (d) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The upper curves show the electropherogram obtained from normal DNA from leukocytes (N), and the lower curves show the electropherogram from tumor DNA (T). In all cases one allele is partially lost in the tumor amplicon.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels—2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH

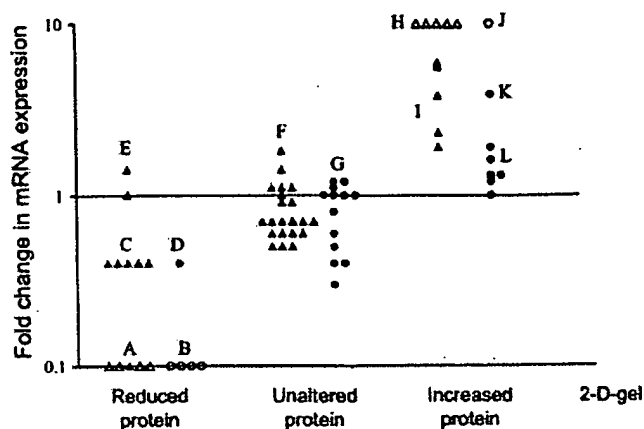


FIG. 4. Correlation between protein levels as judged by 2D-PAGE and transcript ratio. For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (horizontal axis). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (vertical axis). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; Δ, mRNAs that were scored as absent in the invasive tumors (along horizontal axis) or as absent in non-invasive reference (top of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲Δ) were scaled with background suppression, and TCCs 733 and 335 (●○) were scaled without suppression. Both comparisons showed highly significant ($p < 0.005$) differences in mRNA ratios between the groups. Proteins shown were as follows: Group A (from left), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytochrome 15, and cytochrome 13; B (from left), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytochrome 13, and calnexin; C (from left), α -enolase, hnRNP B1, 28-kDa heat shock protein, 14-3-3- ϵ , and pre-mRNA splicing factor; D, mesothelial keratin K7 (type II); E (from top), glutathione S-transferase- π and mesothelial keratin K7 (type II); F (from top and left), adenyl cyclase-associated protein, E-cadherin, keratin 19, calnexin, phosphoglycerate mutase, annexin IV, cytoskeletal γ -actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β -1 subunit; G, (from top and left), TCP20, calnexin, 70-kDa heat shock protein, calnexin, hnRNP H, cytochrome 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione S-transferase- π , and keratin 8; H (from left), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD $^{+}$ -dependent 15 hydroxyprostaglandin dehydrogenase; I (from top), prolyl 4-hydroxylase β -subunit, cytochrome 20, cytochrome 17, prothionin, and fructose 1,6-bisphosphatase; J, annexin II; K, annexin IV; L (from top and left), 90-kDa heat shock protein, prolyl 4-hydroxylase β -subunit, α -enolase, GRP 78, cyclophilin, and cofilin.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

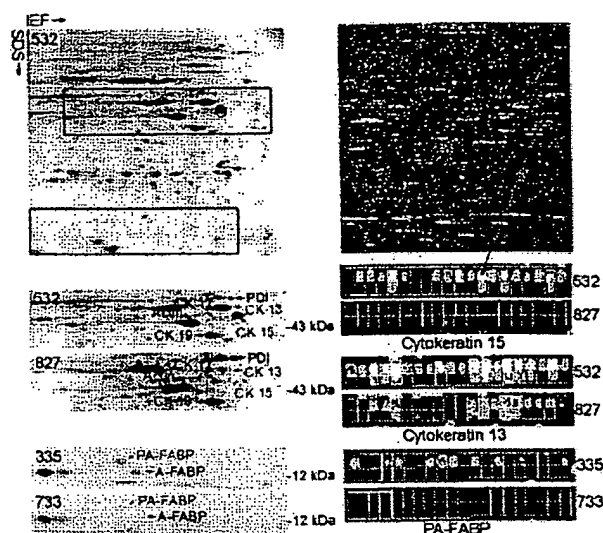


FIG. 5. Comparison of protein and transcript levels in invasive and non-invasive TCCs. The upper part of the figure shows a 2D gel (left) and the oligonucleotide array (right) of TCC 532. The red rectangles on the upper gel highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (red annotation). The tile on the array containing probes for cytokeratin 15 is enlarged below the array (red arrow) from TCC 532 and is compared with TCC 827. The upper row of squares in each tile corresponds to perfect match probes; the lower row corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as black, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the bottom of the figure (left) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the right we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). IEF, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II
Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration ^a	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres ^a	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA)-FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up ^b	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

^a Abs, absent; Pres, present.

^b In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (e.g. in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p–, 9q–, 1q+, Y– (2, 6), and in pT1 tumors, 2q–, 11p–, 11q–, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p– and 9q22–q33– and 9q– and Y–, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14–q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p–, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal

arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicates that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (e.g. signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Nikolaj Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

REFERENCES

- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* 396, 643–649.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46.
- de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* 83, 157–161.
- Brugier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* 12, 1747–1753.
- Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* 50, 1459–1463.
- Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) *c-myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am. J. Pathol.* 146, 1131–1139.
- Richter, J., Jiang, F., Gorog, J. P., Sartorius, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* 57, 2860–2864.
- Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* 153, 1615–1621.
- Bruch, J., Wöhr, G., Hautmann, R., Mattfeldt, T., Brudelein, S., Möller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* 23, 167–174.
- Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* 58, 3555–3560.
- Simon, R., Burger, H., Brinkschmidt, C., Bockel, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* 185, 345–351.
- Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* 110, 87–93.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15, 1359–1367.
- Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* 59, 167–177.
- Celis, J. E., Østergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* 56, 4782–4790.
- Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL.
- Ohlsson, R., Tycko, B., and Sapientza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* 14, 435–438.
- Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* 279, 2118–2121.
- Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* 9, 164–170.
- Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpu, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* 59, 3889–3892.
- Cui, H., Hedborg, F., He, L., Nordenskjöld, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* 57, 4469–4473.
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* 285, 251–254.
- Tsao, J., Yatabe, Y., Markl, I. D., Haiyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* 29, 26–32.
- Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10632–10636.
- Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533–537.
- Ideker, T., Thorsson, V., Raniish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934.

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer^{1,2}

Elizabeth Hyman,³ Päivikki Kauraniemi,³ Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkabloun, Olli-P. Kallioniemi, and Anne Kallioniemi⁴

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E.H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E.H., P.K., S.H., M.W., S.M., E.R., M.R., A.E., O.K., A.K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P.K., A.K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S.H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G.S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O.M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1-6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over

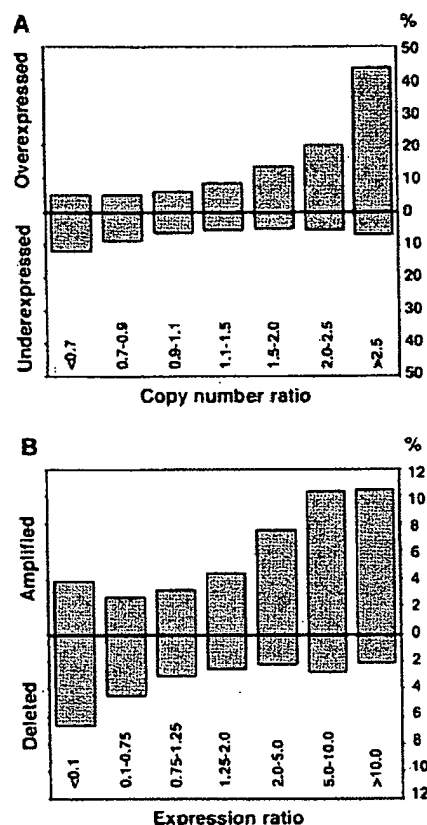


Fig. 1. Impact of gene copy number on global gene expression levels. *A*, percentage of over- and underexpressed genes (*Y* axis) according to copy number ratios (*X* axis). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). *B*, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7 .

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH⁵ (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (a) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (b) identify and characterize those genes whose mRNA expres-

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

² Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org>).

³ Contributed equally to this work.

⁴ To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkkilinjankatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

⁵ The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence in situ hybridization; RT-PCR, reverse transcription-PCR.

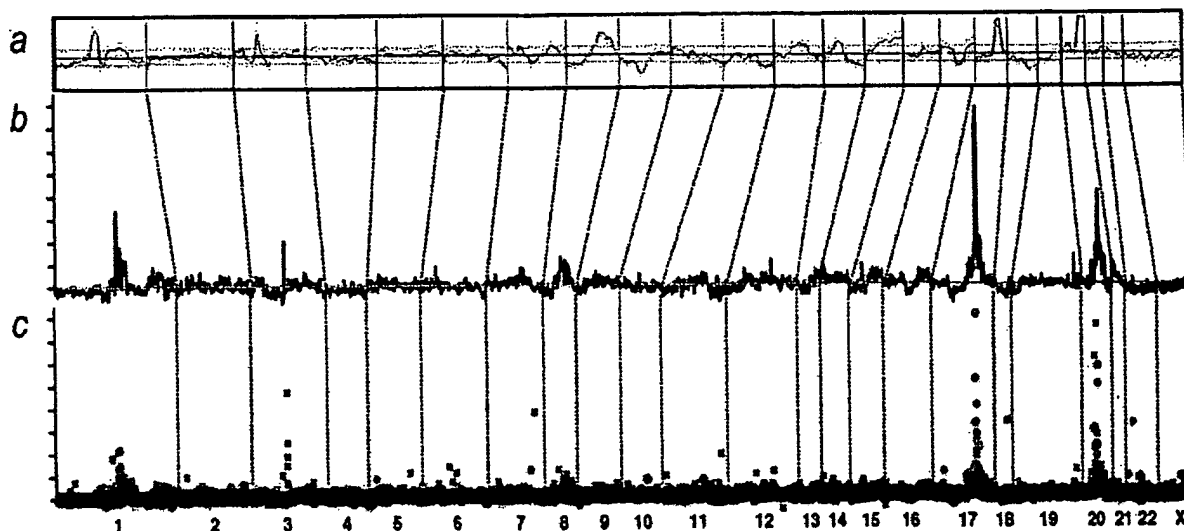


Fig. 2. Genome-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (blue line) across the entire genome from 1p telomere to Xq telomere is shown along with ± 1 SD (orange lines). The black horizontal line indicates a ratio of 1.0; red line, a ratio of 0.8; and green line, a ratio of 1.2. *B-C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. Red horizontal line, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The bright red dots indicate the upper 2%, and dark red dots, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); bright green dots indicate the lowest 2%, and dark green dots, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with black crosses. The chromosome numbers are shown at the bottom of the figure, and chromosome boundaries are indicated with a dashed line.

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μ g of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *A*haI and *R*saI (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μ g of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μ g of reference RNA were labeled with Cy3-dUTP and 3.5 μ g of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units)

were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g , for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1} , σ_{g1} , and m_{g0} , σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α . A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.⁶ A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.⁷ The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

⁶ Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.

⁷ Internet address: www.genome.ucsc.edu.

Table 1 Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1-7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11-8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3-8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22-q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12-q21.2	39.79	42.80	3.0
17q21.32-q21.33	52.47	55.80	3.3
17q22-q23.3	63.81	69.70	5.9
17q23.3-q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12-q13.13	46.45	49.43	3.0
20q13.2-q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5'-GAGCAGAGGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3'.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (*i.e.*, belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal

CGH were validated, with 1q21, 17q12-q21.2, 17q22-q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65-39.25 Mb), and 17q21.3 (52.47-55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22-q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11-p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22-q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

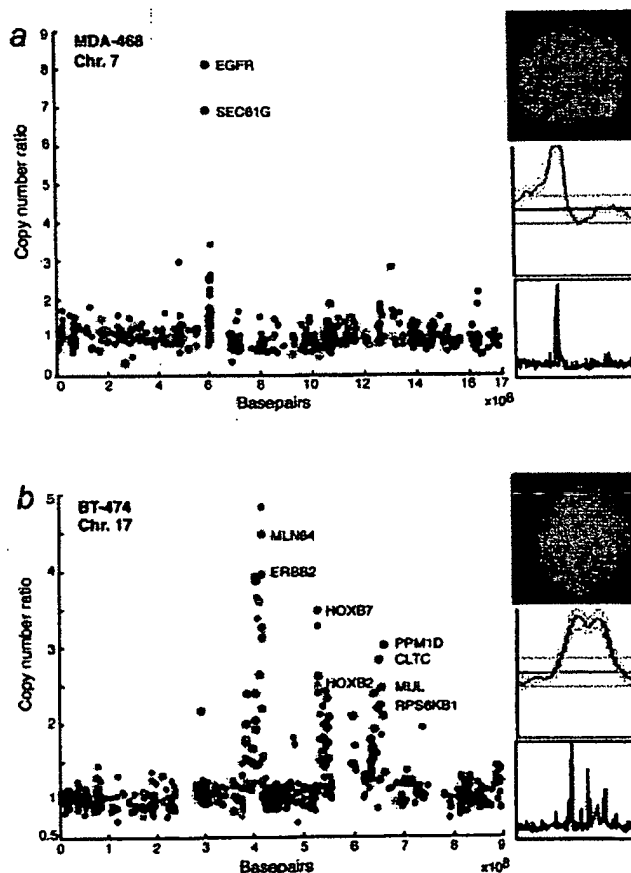


Fig. 3. Annotation of gene expression data on CGH microarray profiles. A, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (red dots) and include the *EGFR* oncogene. B, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (red) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2C. Insets show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (red) and chromosome 7 centromere probe (green) to MDA-468 (A) and *HOXB7*-specific probe (red) and chromosome 17 centromere (green) to BT-474 cells (B).

Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons. Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,⁸ 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

⁹ Internet address: <http://www.ncbi.nlm.nih.gov/enrez>.

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein S6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (a) evidence of a prominent global influence of copy number changes on gene expression levels; (b) a high-resolution map of 24 independent amplicons in breast cancer; and (c) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association

between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (Wash. DC), 286: 531–537, 1999.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (Lond.), 403: 503–511, 2000.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* (Lond.), 406: 536–540, 2000.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. Molecular portraits of human breast tumours. *Nature* (Lond.), 406: 747–752, 2000.
- Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. *Nature* (Lond.), 412: 822–826, 2001.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98: 10869–10874, 2001.
- Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. *Semin. Cancer Biol.*, 9: 125–138, 1999.
- Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. *J. Clin. Oncol.*, 19: 32–40, 2001.
- Knuutila, S., Björkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanski, J., Laramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107–1123, 1998.
- Knuutila, S., Autio, K., and Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.*, 157: 689, 2000.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, 14: 457–460, 1996.
- Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.*, 6: 639–645, 1996.
- Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevas, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. In: F. J. Livesey and S. P. Hunt (eds.), *Functional Genomics*, pp. 113–137. Oxford: Oxford University Press, 2000.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, 23: 41–46, 1999.
- Monni, O., Bärklund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA*, 98: 5711–5716, 2001.
- Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics*, 2: 364–374, 1997.
- Bärklund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Thorst, J., Haas, P., Bucher, C., Sauter, G., et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.*, 92: 1252–1259, 2000.
- Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. *Cytometry*, 45: 83–86, 2001.
- Kauraniemi, P., Bärklund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, 61: 8235–8240, 2001.
- Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. *Genes Chromosomes Cancer*, 34: 104–114, 2002.
- Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kakkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res.*, 62: 2625–2629, 2002.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

- Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.*, 25: 333-337, 2000.
23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc. Natl. Acad. Sci. USA*, 98: 1124-1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linchan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res.*, 61: 8143-8149, 2001.
25. Bärklund, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer*, 20: 372-376, 1997.
26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res.*, 56: 3441-3445, 1996.
27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. *Exp. Cell Res.*, 248: 1-9, 1999.
28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. *J. Cell. Physiol.*, 188: 161-169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol. Cell. Biol.*, 16: 4842-4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the HOXB7 gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. *Oncogene*, 16: 3285-3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. *Cancer Res.*, 61: 6532-6539, 2001.
32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. *Proc. Natl. Acad. Sci. USA*, 98: 4060-4065, 2001.

Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack^{*1‡}, Therese Sørlie[§], Charles M. Perou[¶], Christian A. Rees^{||*}, Stefanie S. Jeffrey^{††}, Per E. Lønning^{**}, Robert Tibshirani^{§§}, David Botstein^{||}, Anne-Lise Børresen-Dale[§], and Patrick O. Brown^{*1¶¶}

Departments of ^{*}Pathology, [†]Genetics, ^{††}Surgery, ^{§§}Health Research and Policy, and ^{¶¶}Biochemistry, and [¶]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; [§]Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; ^{||}Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and ^{††}Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From

this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

Materials and Methods

Tumors and Cell Lines. Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

DNA Labeling and Microarray Hybridizations. Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. “Test” DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The “reference” (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

Data Analysis and Map Positions. Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at <http://rana.lbl.gov>). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

[†]To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: pollack1@stanford.edu.

^{**}Present address: Zyomyx Inc., Hayward, CA 94545.

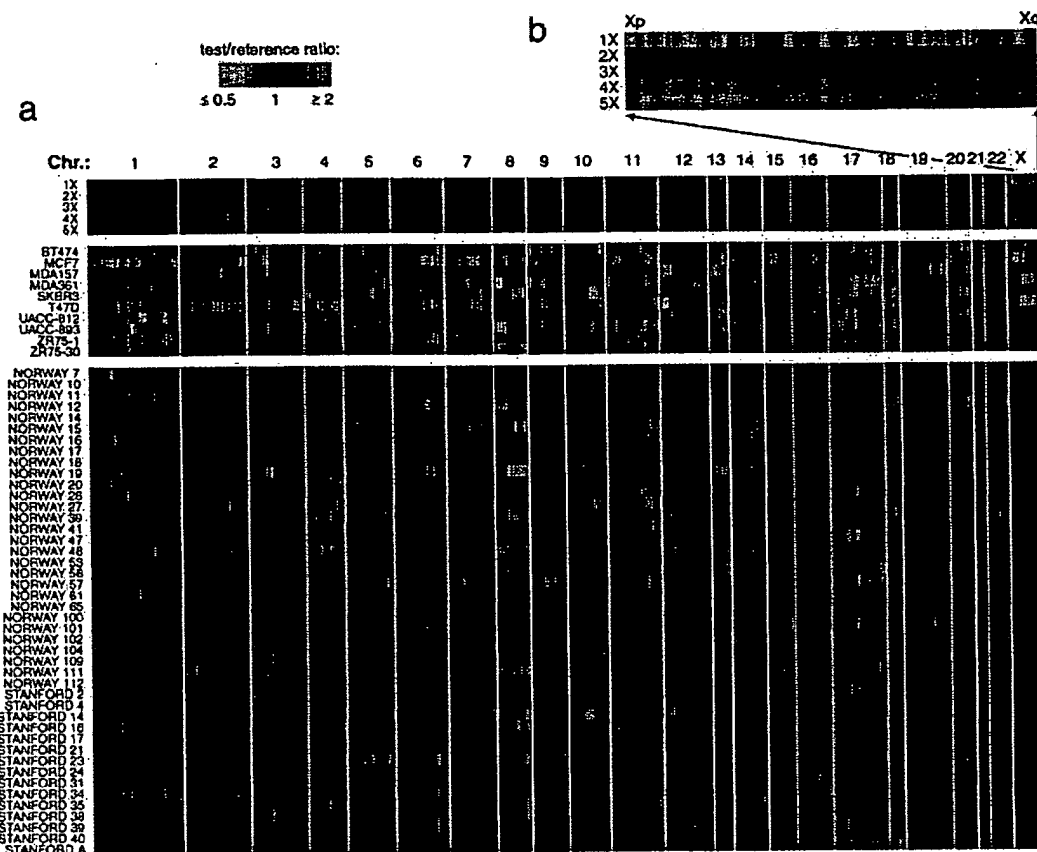


Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a log₂-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (b) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (<http://genome.ucsc.edu/>; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1a; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (<http://genome.ucsc.edu/>) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and

deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1b), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1a), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2–4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total

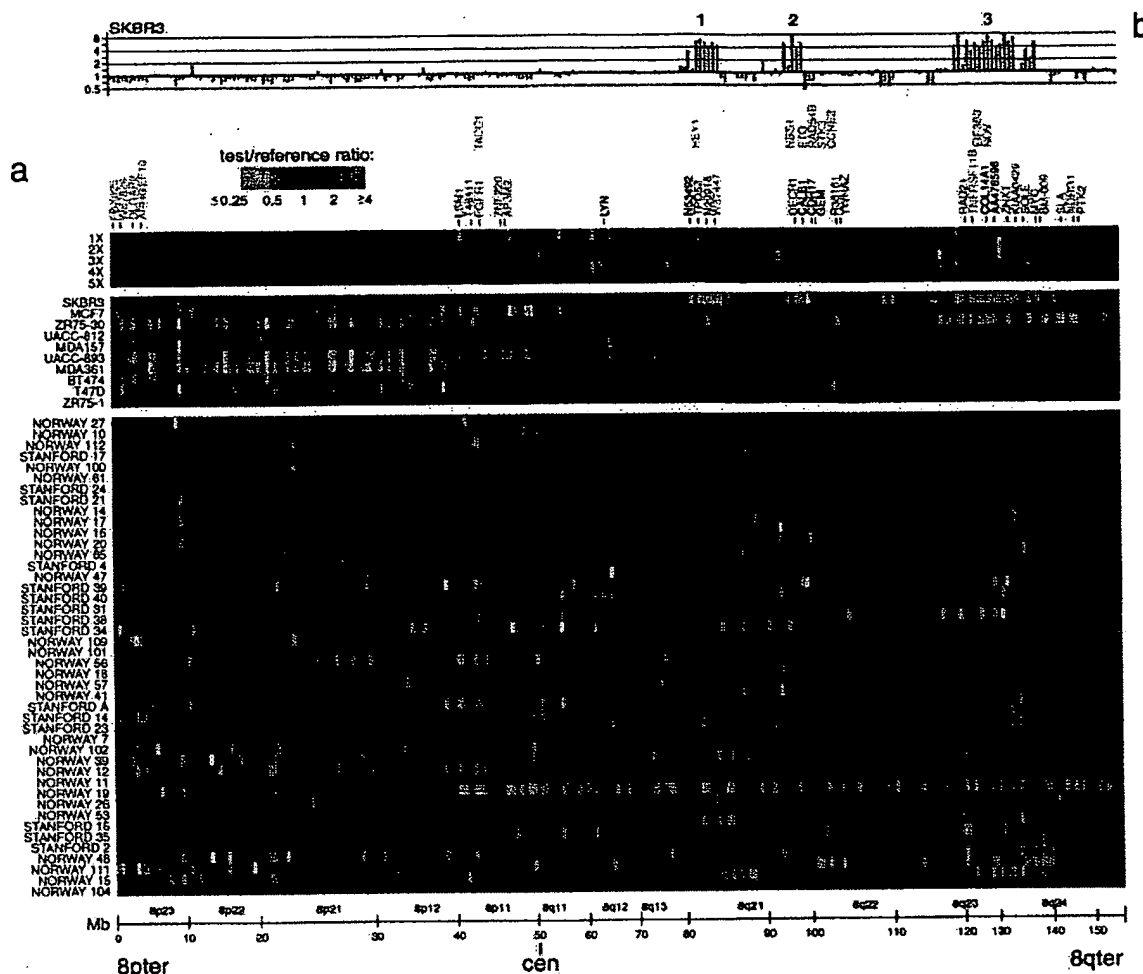


Fig. 2. DNA copy number alteration across chromosome 8 by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a \log_2 pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (b) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a \log_2 scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ($P = 0.008$), consistent with published CGH data (3), estrogen receptor negative ($P = 0.04$), and harboring TP53 mutations ($P = 0.0006$) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2a). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2b). For each of these regions we can define the

boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

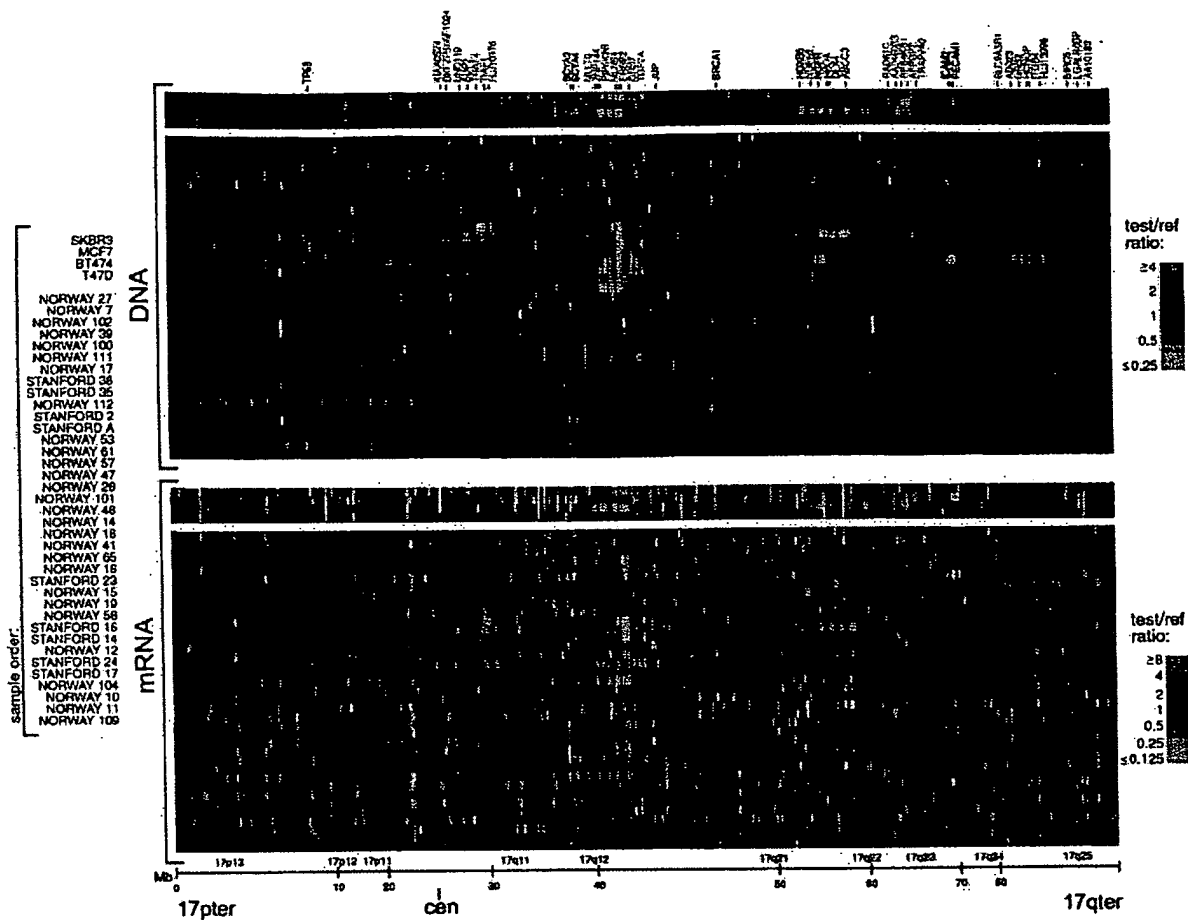


Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (Upper) and mRNA levels (Lower) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (Upper), and the identical sample order is maintained (Lower). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate \log_2 pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4 , and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4a). For both the

breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion (P values for pair-wise Student's t tests comparing adjacent classes: cell lines, 4×10^{-49} , 1×10^{-49} , 5×10^{-5} , 1×10^{-2} ; tumors, 1×10^{-43} , 1×10^{-214} , 5×10^{-41} , 1×10^{-4}). A linear regression of the average $\log(\text{DNA copy number})$, for each class, against average $\log(\text{mRNA level})$ demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4a, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4b). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4c), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4b) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37

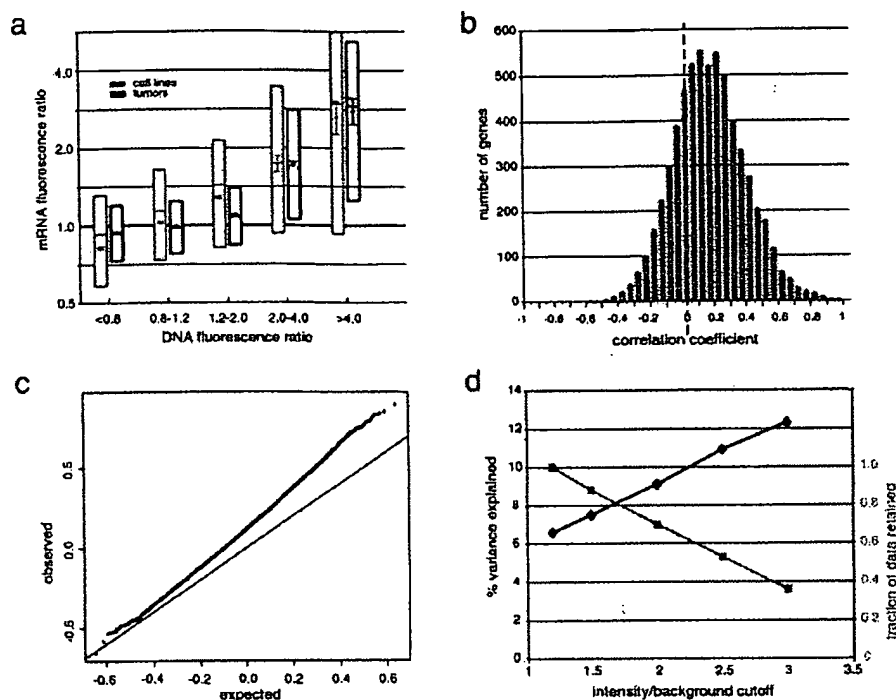


Fig. 4. Genome-wide influence of DNA copy number alterations on mRNA levels. (a) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio (log₂ scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; Y-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. *P* values for pair-wise Student's *t* tests, comparing averages between adjacent classes (moving left to right), are 4×10^{-49} , 1×10^{-49} , 5×10^{-5} , 1×10^{-2} (cell lines), and 1×10^{-43} , 1×10^{-214} , 5×10^{-41} , 1×10^{-4} (tumors). (b) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (c) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (d) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4d). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background >3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4d). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer

cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing

the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stoichiometric relationships in cell metabolism and physiology (e.g., proteosome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

- Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* 258, 818–821.
- Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2156–2160.
- Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* 21, 177–184.
- Forozan, F., Mahlamaki, E. H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* 60, 4519–4525.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* 20, 399–407.
- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* 20, 207–211.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* 23, 41–46.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* 406, 747–752.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98, 10869–10874.
- Schuler, G. D. (1997) *J. Mol. Med.* 75, 694–698.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* 409, 860–921.
- Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* 22, 105–113.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* 25, 333–337.
- Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* 61, 8143–8149.
- Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* 62, 1134–1138.
- Albertson, D. G., Ylstra, B., Seagraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* 25, 144–146.
- Li, R., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 14506–14511.
- Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* 340, 621–630.



TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

HER-2/neu Breast Cancer Predictive Testing

Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.¹ Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.²

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.³

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC; HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in

recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.⁴ HER-2/neu status may be particularly important to establish in women with small (≤ 1 cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.⁵ Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin® (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin® in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin® in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest®) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

CPT code information

HER-2/neu via IHC

88342 (including interpretive report)

HER-2/neu via FISH

- 88271×2 Molecular cytogenetics, DNA probe, each
- 88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells
- 88291 Cytogenetics and molecular cytogenetics, interpretation and report

Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest[®]. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion[™] HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The PathVysion[™] kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1:8-31.
2. "Cancer Rates and Risks", 4th ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3. Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4. Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F. "FISH detection of HER-s/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5. Press, M.F. Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Hung, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

Provided for the clients of

**PATHOLOGY ASSOCIATES MEDICAL LABORATORIES
PACLAB NETWORK LABORATORIES
TRI-CITIES LABORATORY
TREASURE VALLEY LABORATORY**

*For more information, please contact
your local representative.*

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA*†, TODD A. SWANSON*, JAMES W. WELSH*, MARGARET A. ROY‡, DAVID A. LAWRENCE*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE¶, ROBERT L. COHEN*, MONA F. MELHEM**, GENE G. FINLEY**, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,‡‡, AND ARNOLD J. LEVINE§

Departments of *Molecular Oncology, †Molecular Biology, ‡Scientific Computing, and §Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; **University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 β (GSK-3 β) resulting in an increase in β -catenin levels. Stabilized β -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β -catenin levels (9). APC is phosphorylated by GSK-3 β , binds to β -catenin, and facilitates its degradation. Mutations in either APC or β -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- β superfamily, and the homeobox genes, *engrailed*, *gooseoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-06\$2.00/0
PNAS is available online at www.pnas.org.

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

cDNA was synthesized from 2 μ g of poly(A)⁺ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μ g of poly(A)⁺ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

In Situ Hybridization. ³³P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNeasy.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{(\Delta\Delta C_t)}$ where ΔC_t represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The Δ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of $\approx 40,000$ (M_r 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of $\approx 27,000$ (M_r 27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

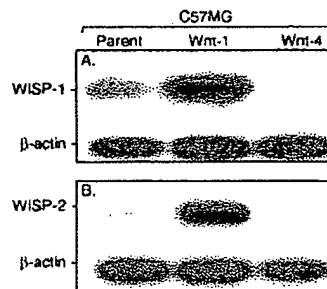


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)⁺ RNA (2 μ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human β -actin probe.

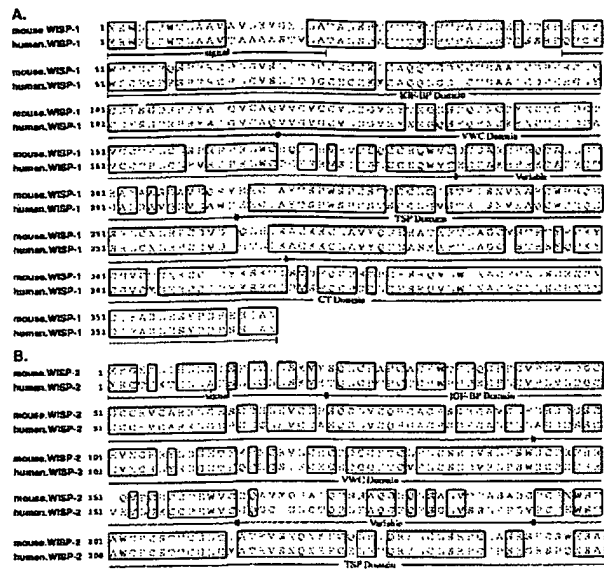


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

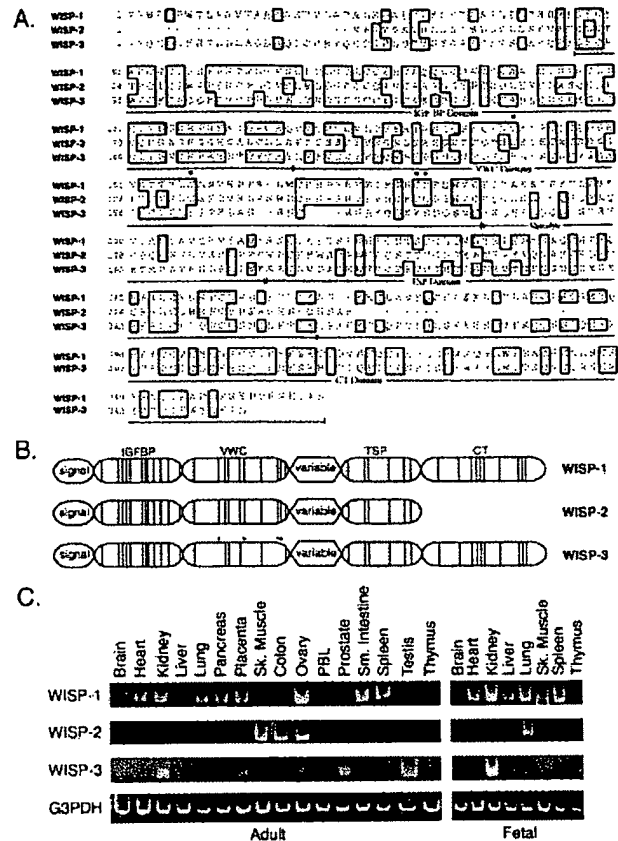


FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSXCSxxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

In Situ Localization of *WISP-1* and *WISP-2*. Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

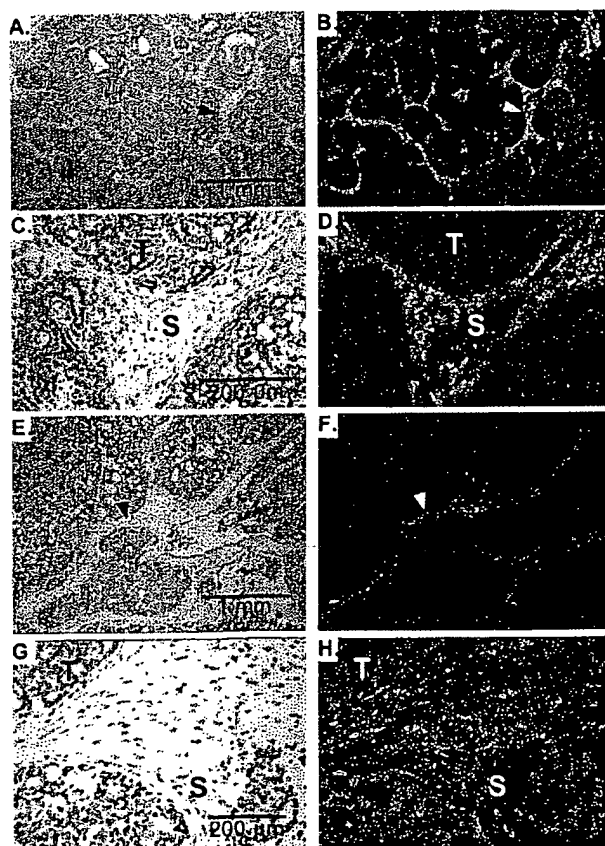


FIG. 4. (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells were negative (G and H).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

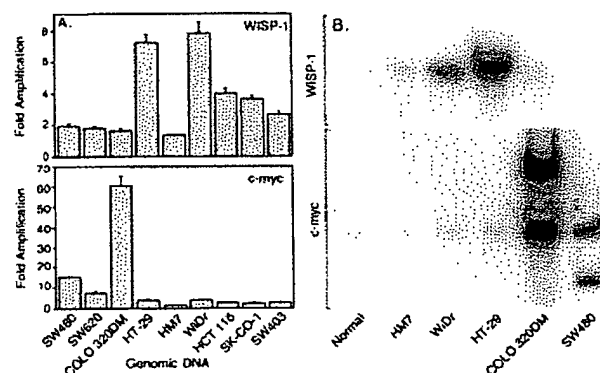


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10 μ g) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

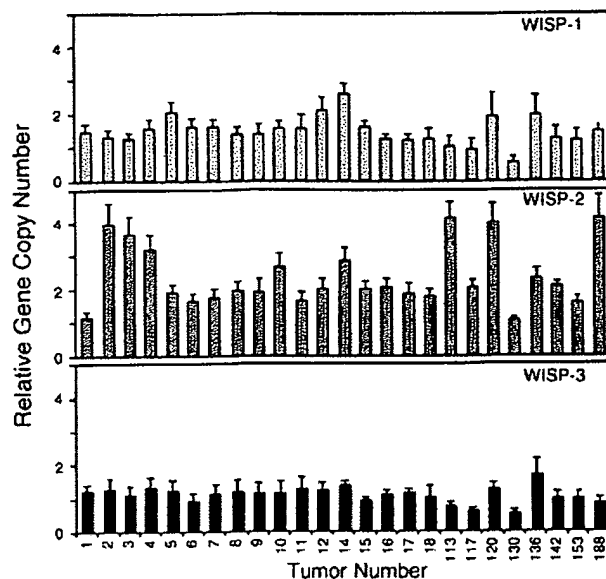


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

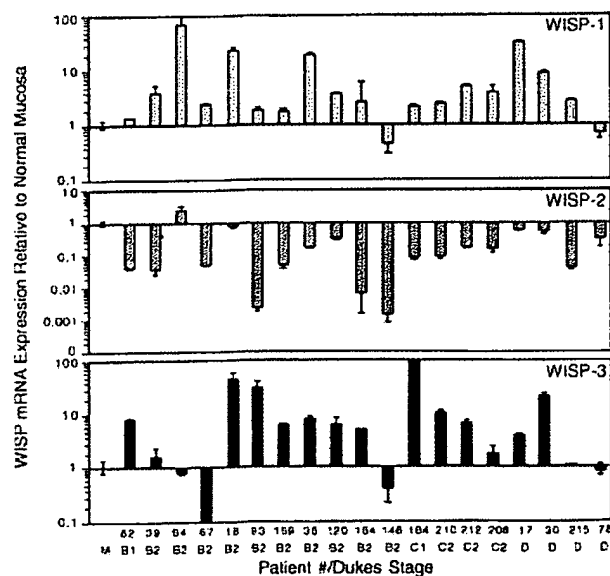


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- β , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_3\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- β 1, which is the stimulus for stromal proliferation (34). TGF- β 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of WISP-1 and WISP-2 in the stromal cells of breast tumors supports this paracrine model.

An analysis of WISP-1 gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of WISP-3 RNA was seen in the absence of DNA amplification. In contrast, WISP-2 DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human WISP-2 was localized to chromosome 20q12-20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36-38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for WISP-2 may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of WISP-2, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which WISP-2 RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of WISP-2 in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the WISP genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the WISPs. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of WISPs as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the WISPs in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

1. Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286-3305.
2. Dale, T. C. (1998) *Biochem. J.* 329, 209-223.
3. Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99-109.
4. van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233-240.
5. Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619-625.
6. Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182-187.
7. Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391-399.

8. Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248-1256.
9. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046-3050.
10. He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509-1512.
11. Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025-6030.
12. Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001-1009.
13. Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278-6286.
14. Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349-1358.
15. Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289-296.
16. Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131-6141.
17. Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171-179.
18. Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326-1334.
19. Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355-6360.
20. Martinier, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729-2732.
21. Bork, P. (1993) *FEBS Lett.* 327, 125-130.
22. Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981-12986.
23. Joliet, V., Martinier, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10-21.
24. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514-19527.
25. Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852-2855.
26. Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195-205.
27. Martinier, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529-2534.
28. Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell. Genet.* 57, 109-111.
29. Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88-94.
30. Garte, S. J. (1993) *Crit. Rev. Oncog.* 4, 435-449.
31. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268-1272.
32. Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269-291.
33. Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090-3096.
34. Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153-161.
35. Wernert, N. (1997) *Virchows Arch.* 430, 433-443.
36. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., *et al.* (1994) *Cancer Res.* 54, 4257-4260.
37. Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187-194.
38. Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., *et al.* (1998) *EMBO J.* 17, 3052-3065.
39. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787-1790.
40. Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169-176.

Variable expression of the translocated *c-abl* oncogene in Philadelphia-chromosome-positive B-lymphoid cell lines from chronic myelogenous leukemia patients

JAMES B. KONOPKA^{*,†}, STEVEN CLARK^{*}, JAMI McLAUGHLIN^{*}, MASAKUZU NITTA[†], YOSHIRO KATO[†], ANNABEL STRIFE[†], BAYARD CLARKSON[†], AND OWEN N. WITTE^{*,§}

^{*}Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90024; and [†]The Laboratory of Hematopoietic Cell Kinetics and The Laboratory of Cancer Genetics and Cytogenetics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

Communicated by Michael Potter, February 10, 1986

ABSTRACT The consistent cytogenetic translocation of chronic myelogenous leukemia (the Philadelphia chromosome, Ph¹) has been observed in cells of multiple hematopoietic lineages. This translocation creates a chimeric gene composed of breakpoint-cluster-region (*bcr*) sequences from chromosome 22 fused to a portion of the *abl* oncogene on chromosome 9. The resulting gene product (P210^{c-abl}) resembles the transforming protein of the Abelson murine leukemia virus in its structure and tyrosine kinase activity. P210^{c-abl} is expressed in Ph¹-positive cell lines of myeloid lineage and in clinical specimens with myeloid predominance. We show here that Epstein-Barr virus-transformed B-lymphocyte lines that retain Ph¹ can express P210^{c-abl}. The level of expression in these B-cell lines is generally lower and more variable than that observed for myeloid lines. Protein expression is not related to amplification of the *abl* gene but to variation in the level of *bcr-abl* mRNA produced from a single Ph¹ template.

Chronic myelogenous leukemia (CML) is a disease of the pluripotent stem cell (1). In greater than 95% of patients, the leukemic cells contain the cytogenetic marker known as the Philadelphia chromosome, or Ph¹ (2). This reciprocal translocation event between the long arms of chromosomes 9 and 22 has been used as a disease-specific marker for diagnosis and evaluation of therapy. Multiple hematopoietic lineages, including myeloid and B-lymphoid, contain Ph¹ in early or chronic phase, as well as in the more acute accelerated and blast crisis phases of the disease.

One molecular consequence of Ph¹ is the translocation of the chromosomal arm containing the *c-abl* gene on chromosome 9 into the middle of the breakpoint-cluster region (*bcr*) gene on chromosome 22 (3-6). Although the precise translocation breakpoints are variable, an RNA-splicing mechanism generates a very similar 8-kilobase (kb) mRNA in each case (5-9). The hybrid *bcr-abl* message encodes a structurally altered form of the *abl* oncogene product, called P210^{c-abl} (10-13), with an amino-terminal segment derived from a portion of the exons of *bcr* on chromosome 22 and a carboxyl-terminal segment derived from a major portion of the exons of the *c-abl* gene on chromosome 9. The chimeric structure of *bcr-abl* and the resulting P210^{c-abl} is similar to the structure of the Abelson murine leukemia virus *gag-abl* genome and resulting P160^{v-abl} transforming gene product. Both proteins have very similar tyrosine kinase activities (10, 11, 14) which can be distinguished by their relative stability to denaturing detergents and by their ATP requirements from the recently described tyrosine kinase activity of the *c-abl* gene product (15).

In concert with structural modification of the amino-terminal portion of the *abl* gene, increased level of expression has been implicated in activation of *c-abl* oncogenic potential. Myeloid and erythroid cell lines and clinical samples derived from acute-phase CML patients contain about 10-fold higher levels of the 8-kb *bcr-abl* mRNA and P210^{c-abl} than the *c-abl* mRNA forms (6 and 7 kb) and P145^{c-abl} gene product (5, 8, 9, 11). The higher level of expression of the chimeric *bcr-abl* message in acute-phase cells is not likely to be solely due to the presence of the *bcr* promoter sequences at the 5' end of the gene, since the normal 4.5-kb and 6.7-kb *bcr*-encoded mRNA species are expressed at an even lower level than the normal *c-abl* messages (5, 6).

We have analyzed a series of Epstein-Barr virus-immortalized B-lymphoid cell lines derived from CML patients (16). With such *in vitro* clonal cell lines, we can evaluate whether the presence of Ph¹ always results in synthesis of the chimeric *bcr-abl* message and protein, and whether the quantitative expression varies for cells of B-lymphoid lineage as compared to previously examined myeloid cell lines. Our results show that cell lines that retain Ph¹ do express *bcr-abl* message and protein, but that the level is generally lower and more variable than previously seen for myeloid cell lines. The demonstration that the Ph¹ chromosomal template can vary in its level of expression of P210^{c-abl} suggests that secondary mechanisms, beyond the translocation itself, contribute to the regulation of the *bcr-abl* gene in different cell types or subclones that derive from the affected stem cell.

MATERIALS AND METHODS

Cells and Cell Labelings. Epstein-Barr virus-transformed B-lymphoid cell lines were established from peripheral blood samples of chronic- and acute-phase CML patients as reported (16). The cell lines are designated according to patient number, karyotype, and lineage. For example, SK-CML7Bt(9,22)-33 refers to CML patient 7, B-lymphoid cell line, 9;22 translocation (Ph¹), cell line 33; and SK-CML7BN-2 refers to B-cell line 2 with a normal karyotype derived from the same patient. Repeat karyotype analysis was performed to verify the retention of Ph¹ just prior to analysis for *abl* protein and RNA. Cells were maintained in RPMI 1640 medium with 20% fetal bovine serum. We have not observed any consistent pattern of *in vitro* growth rate that correlates to the stage of disease at the time of transformation with Epstein-Barr virus. Cells (1.5×10^7) were washed twice with Dulbecco's modified Eagle's medium lacking phosphate and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *bcr*, breakpoint-cluster region; CML, chronic myelogenous leukemia; kb, kilobase(s).

[†]Present address: Department of Genetics, University of Washington, Seattle, WA 98195.

[§]To whom correspondence should be addressed.

supplemented with 5% dialyzed fetal bovine serum. Cells were then resuspended in 2 ml of the minimal medium. Labeling was started with the addition of [32 P]orthophosphate (1 mCi/ml; ICN; 1 Ci = 37 GBq) and continued at 37°C for 3–4 hr.

Immunoprecipitation and Immunoblotting. Immunoprecipitations were carried out as described (10). Cells (1.5×10^7) were washed with phosphate-buffered saline and extracted with 3–5 ml of phosphate lysis buffer (1% Triton X-100/0.1 NaDodSO₄/0.5% deoxycholate/10 mM Na₂HPO₄, pH 7.5/100 mM NaCl) with 5 mM EDTA and 5 mM phenylmethylsulfonyl fluoride. Extracts were clarified by centrifugation and precipitated with normal or rabbit anti-*abl* sera (anti-pEX-2 or anti-pEX-5) (17). The precipitated proteins were electrophoresed in a NaDodSO₄/8% polyacrylamide gel. 32 P-labeled proteins were detected by autoradiography. Alternatively, *abl* proteins were detected by immunoblotting. Extracts from unlabeled cells were clarified, and proteins were concentrated by immunoprecipitation with rabbit antisera against *abl*-encoded proteins [anti-pEX-2 and anti-pEX-5 combined (17)] and then fractionated in 8% acrylamide gels. The proteins were transferred from the gel to nitrocellulose filters, using protease-facilitated transfer (18). The *abl*-encoded proteins were detected using murine monoclonal antibodies as a probe and peroxidase-conjugated goat anti-mouse second stage antibody (Bio-Rad) for development. Rabbit antisera and mouse monoclonal antibodies to *abl* proteins were prepared using bacterially expressed regions of the *v-abl* protein as immunogens (17, 19). Anti-pEX-2 antibodies react with the internal tyrosine kinase domain and anti-pEX-5 antibodies react with the carboxyl-terminal segment of the *abl* proteins.

RNA Analysis. RNA was extracted from 10^8 cells by the NaDodSO₄/urea/phenol method (20). Polyadenylated RNA was purified by oligo(dT) affinity chromatography. Samples were electrophoresed in a 1% agarose/formaldehyde gel and transferred to nitrocellulose. *abl* RNA species were detected by hybridization with a nick-translated *v-abl* fragment probe (21).

DNA Analysis. DNA was prepared from 5×10^7 cells of each cell line and processed for Southern blots with a *v-abl* probe as described (21).

RESULTS

Variable Levels of P210^{c-abl} Are Detected in Ph¹-Positive Cell Lines. Ph¹-positive and Ph¹-negative, Epstein-Barr virus-transformed B-lymphocyte cell lines derived from the same patient were examined for P210^{c-abl} synthesis by immunoprecipitation of [32 P]orthophosphate-labeled cell extracts with anti-*abl* sera (Fig. 1). The normal *c-abl* protein P145^{c-abl} was detected at a similar level in multiple Ph¹-positive and Ph¹-negative cell lines. P210^{c-abl} was only detected in the Ph¹-positive cell lines because the *bcr-abl* chimeric gene which encodes P210^{c-abl} resides on the Ph¹ (4, 5, 11, 13). The level of P210^{c-abl} was about 4- to 5-fold higher than the level of P145^{c-abl} in the SK-CML7Bt-33 cell line (Fig. 1A, +). The Ph¹-positive erythroid-progenitor cell line K562 (C) showed a level of P210^{c-abl} about 10-fold higher than P145^{c-abl}. However, the level of P210^{c-abl} was about one-fifth that of P145^{c-abl} in the Ph¹-positive SK-CML16Bt-1 cell line (Fig. 1B, +). Comparison of different autoradiographic exposures roughly indicated that the level of P210^{c-abl} varies over a 20-fold range between these Ph¹-positive B-cell lines. Analysis of four additional Ph¹-positive B-cell lines demonstrated that the level of P210^{c-abl} fell into two general classes; some cell lines had a level of P210^{c-abl} similar to SK-CML7Bt-33 and others had the low level similar to SK-CML16Bt-1 (Table 1). This differs from previous studies with Ph¹-positive myeloid cell lines and patient samples derived from acute-

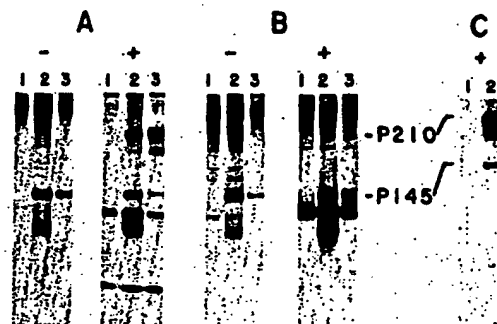


FIG. 1. Detection of variable levels of P210^{c-abl} in Ph¹-positive B-cell lines. Production of P145^{c-abl} and P210^{c-abl} in Epstein-Barr virus-transformed B-cell lines derived from a blast-crisis (A) and a chronic-phase (B) CML patient was examined by metabolic labeling with [32 P]orthophosphate and immunoprecipitation. Ph¹-negative (-) and Ph¹-positive (+) cell lines derived from each patient were analyzed. The Ph¹-negative cell line in A, - is SK-CML7BN-2 and in B, - is SK-CML16BN-1. The Ph¹-positive cell line in A, + is SK-CML7Bt-33 and in B, + is SK-CML16Bt-1. The K562 cell line, a Ph¹-positive erythroid progenitor cell line spontaneously derived from a blast-crisis patient (33), is represented in C. Cells (1.5×10^7) were metabolically labeled with 2 mCi of [32 P]orthophosphate for 3–4 hr and then were extracted and clarified by centrifugation. Samples were immunoprecipitated with control normal serum (lanes 1), anti-pEX-2 (lanes 2), or anti-pEX-5 (lanes 3) and analyzed by NaDodSO₄/8% PAGE followed by autoradiography with an intensifying screen (3 days for A and C, 10 days for B).

phase CML patients, in which P210^{c-abl} was detected at a 10-fold higher level than P145^{c-abl} (refs. 10 and 11; Table 1). There was no large difference in level of chimeric mRNA and P210^{c-abl} expressed in four myeloid/erythroid-lineage Ph¹-positive cell lines (K562, EM2, EM3, CML22, and BV173; refs. 9 and 11), despite a 4- to 5-fold amplification of *abl*-related sequences in the K562 cell line.

Detection of different levels of P210^{c-abl} in Fig. 1 could be due to decreased phosphorylation of P210^{c-abl}, a lower level of P210^{c-abl} synthesis, or altered stability of the protein. To help distinguish among these possibilities, the steady-state level of P210^{c-abl} in the cell lines was assayed by immunoblotting. The results show that SK-CML7Bt-33 (Fig. 2A, +) had a higher level of P210^{c-abl} than P145, similar to the results with metabolic labeling (Fig. 1). We did not detect P210^{c-abl} by immunoblotting with 2×10^7 cells of line SK-CML8Bt-3 (Fig. 2B, +). Reconstruction experiments using dilutions of cell extracts showed that we could detect about 5–10% the level of P210^{c-abl} expressed in the K562 cell line (data not shown). We infer that the steady-state level of P210^{c-abl} in SK-CML8Bt-3 is lower than the level in SK-CML7Bt-33 by a factor of at least 10. The level of P210^{c-abl} detected in these assays correlated with the amount of P210^{c-abl} tyrosine kinase activity that could be detected *in vitro* (data not shown).

Different Levels of P210^{c-abl} Are Reflected in the Amount of Stable *bcr-abl* mRNA. To identify the basis for detection of variable levels of P210^{c-abl}, we examined the production of the *abl* RNA. RNA blot hybridization analysis using a *v-abl* probe (Fig. 3) showed that the normal 6- and 7-kb *c-abl* mRNAs were present at a similar level in Ph¹-positive and -negative cell lines derived from different patients. However, the 8-kb mRNA that encodes P210^{c-abl} was detected at a 10-fold higher level in SK-CML7Bt-33 (Fig. 3A, +) than in SK-CML16Bt-1 (B, +), which correlated with the relative level of P210^{c-abl} detected in each cell line. Analysis of additional cell lines demonstrated that the level of 8-kb RNA directly correlated with the level of P210^{c-abl} (Table 1). The variation in level of 8-kb RNA detected in these cell lines was not due to loss or gain of Ph¹, because cytogenetic analysis confirmed the presence of Ph¹ in these cell lines (ref. 16 and

Table 1. Relative levels of *bcr-abl* expression in Epstein-Barr virus-immortalized B-cell lines and myeloid CML lines

Cell line*	CML phase†	Ph ⁺	P210‡	8-kb mRNA§
SK-CML7BN-2	BC	-	-	-
SK-CML8BN-10	Chronic	-	-	-
SK-CML8BN-12	Chronic	-	-	-
SK-CML16BN-1	Chronic	-	-	-
SK-CML35BN-1	Chronic	-	-	-
SK-CML7B5-33	BC	+	+++	+++
SK-CML21Bt-1	Acc	+	+++	+++
SK-CML21Bt-6	Acc	+	+++	+++
SK-CML8Bt-3	Chronic	+	+	±
SK-CML16Bt-1	Chronic	+	+	+
SK-CML35Bt-2	Chronic	+	+	+
K562	BC	+	+++++	+++++
BV173	BC	+	+++++	+++++
EM2	BC	+	+++++	+++++

*Cell lines derived from CML patients by transformation with Epstein-Barr virus as described (16). Names of cell lines indicate patient number and Ph⁺ status: SK-CML7Bt indicates a cell line derived from patient 7 that carries the 9;22 Ph⁺ translocation; N indicates a normal karyotype. Myeloid-erythroid cell lines (K562, EM2, and BV173) are described in previous publications (9, 11, 22, 33).

†Status of patient at the time cell line was derived. BC, blast crisis; Acc, accelerated phase.

‡Presence (+) or absence (-) of Ph⁺ as demonstrated by karyotypic or Southern blot analysis.

§P210^{abl} detected as described in legend to Fig. 1. B-cell lines derived from blast-crisis and accelerated-phase patients had levels of P210 3- to 5-fold higher (++++) than levels of P145. Chronic-phase-derived cell lines had P210 levels lower than or just equivalent (+) to the level of P145. Myeloid and erythroid lines had levels of P210 5- to 10-fold higher than P145 (+++++).

¶Eight-kilobase *bcr-abl* mRNA detected as described in legend to Fig. 2. Symbols: ±, borderline detectable; +++++, level of 8-kb mRNA 5- to 10-fold higher than that of the 6- and 7-kb *c-abl* mRNA species; +++, level of 8-kb mRNA 3- to 5-fold higher than that of the 6- and 7-kb species; +, a level approximately equivalent to that of the 6- and 7-kb messages.

data not shown). There was no difference in the copy number of *abl*-related sequences as judged by Southern blot analysis (Fig. 4). Only the K562 cell line control showed an amplification of *abl* sequences, as previously reported (22, 23). These combined data suggest that differential *bcr-abl* mRNA expression from a single gene template is responsible for the variable levels of P210^{abl} detected. This could be mediated

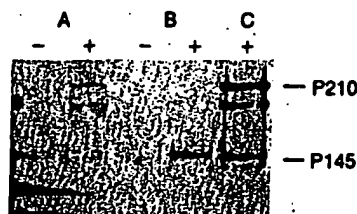


FIG. 2. Analysis of steady-state *abl* protein levels by immunoblotting. Cell extracts prepared from 2×10^7 cells of lines SK-CML7BN-2 (A, -), SK-CML7Bt-33 (A, +), SK-CML8BN-10 (B, -), and SK-CML8Bt-3 (B, +) were concentrated by immunoprecipitation with anti-pEX-2 plus anti-pEX-5. Samples were then electrophoresed in a NaDodSO₄/8% polyacrylamide gel and transferred to nitrocellulose, using protease-facilitated transfer (18). *abl* proteins were detected using a mixture of two monoclonal antibodies directed against the pEX-2 and pEX-5 *abl*-protein fragments produced in bacteria (19) as a probe and a peroxidase-conjugated goat anti-mouse second-stage antibody (Bio-Rad) for development.

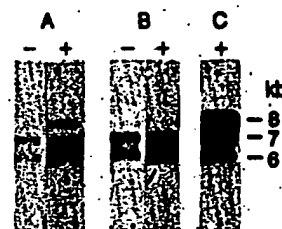


FIG. 3. Comparison of *abl* RNA levels in Ph⁺-positive and -negative B-cell lines. The levels of the normal 6- and 7-kb *c-abl* RNAs and the 8-kb *bcr-abl* RNA were analyzed by blot hybridization using a *v-abl* probe. RNA was extracted from Ph⁺-negative lines SK-CML7BN-2 (A, -) and SK-CML16BN-1 (B, -), from Ph⁺-positive lines SK-CML6Bt-33 (A, +) and SK-CML16Bt-3 (B, +), and from line K562 (C, +) by the NaDodSO₄/urea/phenol method (20). Polyadenylated RNA was purified by oligo(dT) affinity chromatography, and 15 µg of each sample was electrophoresed in a 1% agarose/formaldehyde gel and then transferred to nitrocellulose. The blotted RNAs were hybridized with a nick-translated *v-abl* fragment probe (21) and then autoradiographed for 4 days.

by factors influencing the transcription rate of the *bcr-abl* gene or the stability of the mRNA.

DISCUSSION

Several lines of evidence suggest that formation of Ph⁺ is not the primary event that affects the stem cell in CML. Patients have been identified that present with the clinical picture of CML but only later develop Ph⁺ (1). This observation, coupled with studies of *G6PD* (glucose-6-phosphate dehydrogenase)-heterozygous females with CML that demonstrate stem-cell clonality by isozyme analysis among cell

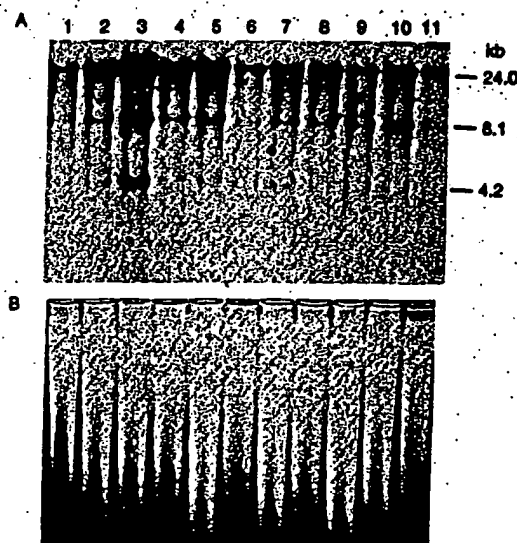


FIG. 4. Southern blot analysis of *abl* sequences in Ph⁺-positive and -negative B-cell lines. High molecular weight DNA (15 µg) was digested with restriction endonuclease *Bam*HI, separated in a 0.8% agarose gel, and then transferred to nitrocellulose. The blotted DNA fragments were hybridized with a nick-translated, 2.4-kb *Bgl*II *v-abl* fragment (1.5×10^6 cpm/µg; ref. 21) and exposed for 4 days. (A) Autoradiogram of *abl*-specific fragments in cell lines HL-60 (lane 1), EM2 (lane 2), K562 (lane 3), SK-CML7Bt-33 (lane 4), SK-CML8Bt-3 (lane 5), SK-CML16Bt-1 (lane 6), SK-CML21Bt-6 (lane 7), SK-CML35Bt-2 (lane 8), SK-CML7BN-2 (lane 9), SK-CML8BN-2 (lane 10), and SK-CML35BN-1 (lane 11). (B) Ethidium bromide staining of agarose gel prior to transfer to nitrocellulose, showing the level of variation in amount of DNA loaded per lane.

populations that lack the Ph¹ marker, supports a secondary or complementary role for Ph¹ in the progression of the disease (24, 25). This chromosome marker is found in chronic, accelerated, and blast-crisis phases of the disease. It is likely that Ph¹ confers some growth advantage, since cells with the marker chromosome eventually predominate the marrow and peripheral blood even in chronic phase. During the phase of blast crisis, many patients develop additional chromosome abnormalities, including duplication of Ph¹, a variety of trisomies, and complex translocations (26). This is suggestive evidence for Ph¹ being a necessary but not sufficient genetic change for the full evolution of the disease.

The realization that one molecular result of Ph¹ is the generation of a chimeric *bcr-abl* protein with functional characteristics and structure analogous to the *gag-abl* transforming protein of the Abelson murine leukemia virus strengthens the argument for an important role of Ph¹ in the pathogenesis of CML. Although the Abelson virus is generally considered a rapidly transforming retrovirus, its effects can range from overcoming growth factor requirements, to cellular lethality, to induction of highly oncogenic tumors in a number of hematopoietic cell lineages (27, 28). Even in the transformation of murine cell targets, there are several lines of evidence that suggest that the growth-promoting activity of the *v-abl* gene product is complemented by further cellular changes in the production of the malignant-cell phenotype (29-31).

The regulation of *bcr-abl* gene expression is complex because the 5' end of the gene is derived from the non-*abl* sequences, *bcr*, normally found on chromosome 22 (6). The level of stable message for the normal *bcr* gene and the normal *abl* gene are both much lower than the level of the *bcr-abl* message and protein from cell lines and clinical specimens derived from myeloid blast-crisis patients (5, 6, 11). Therefore, the high level of *bcr-abl* expression cannot simply be attributed to the regulatory sequences associated with *bcr*. Possibly, creation of the chimeric gene disrupts the normal regulatory sequences and results in a higher level of expression. Variation in *bcr-abl* expression may result from secondary changes in the structure of the chimeric gene or function of *trans*-acting factors that occur during evolution of the disease. Our analysis of P210^{c-abl} and the 8-kb mRNA in Epstein-Barr virus-transformed Ph¹-positive B-cell lines demonstrates that stable message and protein levels from the *bcr-abl* gene can vary over a wide range. This variation does not result from a change in the number of *bcr-abl* templates secondary to gene amplification but more likely from changes in either transcription rate or mRNA stability. We suspect this range of *bcr-abl* expression is not limited to lymphoid cells. Analysis of peripheral blood leukocytes derived from an unusual CML patient who has been in chronic phase with myeloid predominance for 16 years showed a level of P210^{c-abl} one-fifth that of P145^{c-abl}, as detected by metabolic labeling with [³²P]orthophosphate and immunoprecipitation (S.C., O.N.W., and P. Greenberg, unpublished observations). Lower levels of expression of the chimeric mRNA have been demonstrated in clinical samples from chronic-phase CML patients compared to acute-phase CML patients (9). Others have reported chronic-phase patients with variable but, in some cases, relatively high levels of the *bcr-abl* mRNA (32). The sampling variation and the heterogeneous mixture of cell types in clinical samples complicate such analyses. Further work is needed to evaluate whether there is a defined change in P210^{c-abl} expression during the progression of CML. It is interesting to note that among the limited sample of Ph¹-positive B-cell lines we have examined (Table 1), we have seen higher levels of P210^{c-abl} in those derived from patients at more advanced stages of the disease.

It will be important to search for cell-type-specific mechanisms that might regulate expression of *bcr-abl* from Ph¹.

We thank Bonnie Hechinger and Carol Crookshank for excellent secretarial assistance and Margaret Newman for excellent technical assistance. This work was supported by grants from the National Institutes of Health (to O.N.W. and B.C.). J.B.K. was supported as a predoctoral fellow on the Public Health Service Cellular and Molecular Biology Training Grant GM07185. S.C. is a postdoctoral fellow of the Leukemia Society of America.

1. Champlin, R. E. & Golde, D. W. (1985) *Blood* 65, 1039-1047.
2. Rowley, J. D. (1973) *Nature (London)* 243, 290-291.
3. Heisterkamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., de Klein, A., Bartram, C. R. & Grosfeld, G. (1983) *Nature (London)* 306, 239-242.
4. Bartram, C. R., de Klein, A., Hagemeijer, A., van Agthoven, T., van Kessel, A. G., Bootsma, D., Grosfeld, G., Ferguson-Smith, M. A., Davies, T., Stone, M., Heisterkamp, N., Stephenson, J. R. & Groffen, J. (1983) *Nature (London)* 306, 277-280.
5. Shivelman, E., Lifshitz, B., Gale, R. P. & Canaani, D. (1985) *Nature (London)* 315, 550-554.
6. Heisterkamp, N., Stam, K. & Groffen, J. (1985) *Nature (London)* 315, 758-761.
7. Groffen, J., Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C. R. & Grosfeld, G. (1984) *Cell* 36, 93-99.
8. Gale, R. P. & Canaani, E. (1984) *Proc. Natl. Acad. Sci. USA* 81, 5648-5652.
9. Collins, S., Kubonishi, L., Miyoshi, I. & Groudine, M. T. (1984) *Science* 225, 72-74.
10. Konopka, J. B., Watanabe, S. M. & Witte, O. N. (1984) *Cell* 7, 1035-1042.
11. Konopka, J. B., Watanabe, S. M., Singer, J., Collins, S. & Witte, O. N. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1810-1814.
12. Kloetzer, W., Kurzrock, R., Smith, L., Talpaz, M., Spiller, M., Guterman, J. & Arlinghaus, R. (1985) *Virology* 140, 230-238.
13. Kozbor, D., Giallongo, A., Sierzega, M. E., Konopka, J. B., Witte, O. N., Showe, L. C. & Croce, C. M. (1985) *Nature (London)*, in press.
14. Davis, R. L., Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell Biol.* 5, 204-213.
15. Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell Biol.* 5, 3116-3123.
16. Nitta, M., Kato, Y., Strife, A., Wachter, M., Fried, J., Perez, A., Jhanwar, S., Duigou, R., Chaganti, R. S. K. & Clarkson, B. (1985) *Blood* 66, 1053-1061.
17. Konopka, J. B., Davis, J. L., Watanabe, S. M., Ponticelli, A. S., Schiff-Maker, L., Rosenberg, N. & Witte, O. N. (1984) *Virology* 51, 223-232.
18. Gibson, W. (1981) *Anal. Biochem.* 118, 1-3.
19. Schiff-Maker, L., Konopka, J. B., Clark, S., Witte, O. N. & Rosenberg, N. (1986) *J. Virol.* 57, 1182-1186.
20. Schwartz, R. C., Sonenshein, G. E., Bothwell, A. & Gelfand, M. L. (1981) *J. Immunol.* 126, 2104-2108.
21. Goff, S. P., Gilboa, E., Witte, O. N. & Baltimore, D. (1980) *Cell* 22, 777-785.
22. Collins, S. J. & Groudine, M. T. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4813-4817.
23. Selden, J. R., Emanuel, B. S., Wang, E., Cannizzaro, L., Palumbo, A., Erikson, J., Nowell, P. C., Rovera, G. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7289-7292.
24. Fialkow, P. J., Martin, P. J., Najfeld, V., Penfold, G. K., Jacobson, R. J. & Hansen, J. A. (1981) *Blood* 58, 158-163.
25. Martin, P. J., Najfeld, V. & Fialkow, P. J. (1982) *Can. Gen. Cytogenet.* 6, 359-368.
26. Rowley, J. D. (1980) *Annu. Rev. Genet.* 14, 17-40.
27. Whitlock, C. A. & Witte, O. N. (1984) *Adv. Immunol.* 37, 74-98.
28. Pierce, J. H., Di Fiore, P. P., Aaronson, S. A., Potter, M., Pumphrey, J., Scott, A. & Ihle, J. N. (1985) *Cell* 41, 685-693.
29. Whitlock, C. A., Ziegler, S. & Witte, O. N. (1983) *Mol. Cell Biol.* 3, 596-604.
30. Wolf, D., Harris, N. & Rotter, V. (1984) *Cell* 38, 119-126.
31. Klein, G. & Klein, E. (1985) *Nature (London)* 315, 190-195.
32. Stam, K., Jr., Heisterkamp, N., Grosfeld, G., de Klein, A., Verma, R., Coleman, M., Dosik, H. & Groffen, J. (1985) *N. Engl. J. Med.* 313, 1429-1433.
33. Lozzio, C. B. & Lozzio, B. B. (1975) *Blood* 45, 321-334.

Aneuploidy and cancer

Subrata Sen, PhD

Numeric aberrations in chromosomes, referred to as aneuploidy, is commonly observed in human cancer. Whether aneuploidy is a cause or consequence of cancer has long been debated. Three lines of evidence now make a compelling case for aneuploidy being a discrete chromosome mutation event that contributes to malignant transformation and progression process. First, precise assay of chromosome aneuploidy in several primary tumors with *in situ* hybridization and comparative genomic hybridization techniques have revealed that specific chromosome aneusomies correlate with distinct tumor phenotypes. Second, aneuploid tumor cell lines and *in vitro* transformed rodent cells have been reported to display an elevated rate of chromosome instability, thereby indicating that aneuploidy is a dynamic chromosome mutation event associated with transformation of cells. Third, and most important, a number of mitotic genes regulating chromosome segregation have been found mutated in human cancer cells, implicating such mutations in induction of aneuploidy in tumors. Some of these gene mutations, possibly allowing unequal segregations of chromosomes, also cause tumorigenic transformation of cells *in vitro*. In this review, the recent publications investigating aneuploidy in human cancers, rate of chromosome instability in aneuploidy tumor cells, and genes implicated in regulating chromosome segregation found mutated in cancer cells are discussed. Curr Opin Oncol 2000, 12:82-88 © 2000 Lippincott Williams & Wilkins, Inc.

The University of Texas, M.D. Anderson Cancer Center, Department of Laboratory Medicine, Houston, Texas, USA

Correspondence to Subrata Sen, PhD, The University of Texas, MD Anderson Cancer Center, Department of Laboratory Medicine, Box 054, 1515 Holcombe Blvd., Houston, TX 77030, USA; tel: 713-792-2560; fax: 713-792-4094; e-mail: ssen@mdanderson.org

Current Opinion in Oncology 2000 12:82-88

Abbreviations

CGH comparative genomic hybridization
CHE Chinese hamster embryo cells
FISH fluorescence in situ hybridization
HPRC hereditary papillary renal carcinoma
ISH *in situ* hybridization

ISSN 1040-8746 © 2000 Lippincott Williams & Wilkins, Inc.

Cancer research over the past decade has firmly established that malignant cells accumulate a large number of genetic mutations that affect differentiation, proliferation, and cell death processes. In addition, it is also recognized that most cancers are clonal, although they display extensive heterogeneity with respect to karyotypes and phenotypes of individual clonal populations. It is estimated that numeric chromosomal imbalance, referred to as *aneuploidy*, is the most prevalent genetic change recorded among over 20,000 solid tumors analyzed thus far [1]. Phenotypic diversity of the clonal populations in individual tumors involve differences in morphology, proliferative properties, antigen expression, drug sensitivity, and metastatic potentials. It has been proposed that an underlying acquired genetic instability is responsible for the multiple mutations detected in cancer cells that lead to tumor heterogeneity and progression [2]. In a somewhat contradictory argument, it has also been suggested that clonal expansion due to selection of cells undergoing normal rates of mutation can explain malignant transformation and progression process in humans [3]. Acquired genetic instability, nonetheless, is considered important for more rapid progression of the disease [4••]. Although the original hypothesis on genetic instability in cancer primarily focused on chromosome imbalances in the form of aneuploidy in tumor cells, the actual relevance of such mutations in cancer remains a controversial issue.

Whether or not aneuploidy contributes to the malignant transformation and progression process has long been debated. A prevalent idea on genetics of cancer referred to as "somatic gene mutation hypothesis" contends that gene mutations at the nucleotide level alone can cause cancer by either activating cellular proto-oncogenes to dominant cancer causing oncogenes and/or by inactivating growth inhibitory tumor suppressor genes. In this scheme of things chromosomal instability in the form of aneuploidy is a mere consequence rather than a cause of malignant transformation and progression process.

In this review, some of the recent observations on the subject are discussed and compelling evidence is provided to suggest that aneuploidy is a distinct form of genetic instability in cancer that frequently correlates with specific phenotypes and stages of the disease. Furthermore, discrete genetic targets affecting chromosomal stability in cancer cells, recently identified, are also discussed. These data provide a new direction toward elucidating the molecular mechanisms responsi-

ble for induction of aneuploidy in cancer and may eventually be exploited as novel therapeutic targets in the future.

Genetic alterations in cancer

Alterations in many genetic loci regulating growth, senescence, and apoptosis, identified in tumor cells, have led to the current understanding of cancer as a genetic disease. The genetic changes identified in tumors include: subtle mutations in genes at the nucleotide level; chromosomal translocations leading to structural rearrangements in genes; and numeric changes in either partial segments of chromosomes or whole chromosomes (aneuploidy) causing imbalance in gene dosage.

For the purpose of this review, both segmental and whole chromosome imbalances leading to altered DNA dosage in cancer cells are included as examples of aneuploidy.

Incidence of aneuploidy in cancer

Evidence of aneuploidy involving one or more chromosomes have been commonly reported in human tumors. Although these observations were initially made using classic cytogenetic techniques late in a tumor's evolution and were difficult to correlate with cancer progression, more recent studies have reported association of specific nonrandom chromosome aneuploidy with different biologic properties such as loss of hormone dependence and metastatic potential [5].

Classic cytogenetic studies performed on tumor cells had serious limitations in scope because they were applicable only to those cases in which mitotic chromosomes could be obtained. Because of low spontaneous rates of cell division in primary tumors, analyses depended on cells either derived selectively from advanced metastases or those grown *in vitro* for variable periods of time. In both instances, metaphases analyzed represented only a subset of primary tumor cell population. Two major advances in cytogenetic analytic techniques, *in situ* hybridization (ISH) and comparative genomic hybridization (CGH), have allowed better resolution of chromosomal aberrations in freshly isolated tumor cells [6]. ISH analyses with chromosome-specific DNA probes, a powerful adjunct to metaphasic analysis, allows assessment of chromosomal anomalies within tumor cell populations in the contexts of whole nuclear architecture and tissue organization. CGH allows genome wide screening of chromosomal anomalies without the use of specific probes even in the absence of prior knowledge of chromosomes involved. Although both techniques have certain limitations in terms of their resolution power, they nonetheless provide a better approximation of chromosomal changes occurring among tumors of various histology, grade, and stage

compared with what was possible with the classic cytogenetic techniques. Genomic ploidy measurements have also been performed at the DNA level with flow cytometry and cytofluorometric methods. Although these assays underestimate chromosome ploidy due to a chromosomal gain occasionally masking a chromosomal loss in the same cell, several studies using these methods have supported the conclusion that DNA aneuploidy closely associates with poor prognosis in various cancers [7,8]. This discussion of some recent examples published on aneuploidy in cancer includes discussion of studies dealing with DNA ploidy measurements as well. Most of these observations are correlative without direct proof of specific involvement of genes on the respective chromosomes. Identification of putative oncogenes and tumor suppressor genes on gained and lost chromosomes in aneuploid tumors, however, are providing strong evidence that chromosomes involved in aneuploidy play a critical role in the tumorigenic process.

In renal tumors, either segmental or whole chromosome aneuploidy appears to be uniquely associated with specific histologic subtypes [9]. Tumors from patients with hereditary papillary renal carcinomas (HPRC) commonly show trisomy of chromosome 7, when analyzed by CGH. Germline mutations of a putative oncogene *MET* have been detected in patients with HPRC. A recent study [10] has demonstrated that an extra copy of chromosome 7 results in nonrandom duplication of the mutant *MET* allele in HPRC, thereby implicating this trisomy in tumorigenesis. The study suggested that mutation of *MET* may render the cells more susceptible to errors in chromosome replication, and that clonal expansion of cells harboring duplicated chromosome 7 reflects their proliferative advantage. In addition to chromosome 7, trisomy of chromosome 17 in papillary tumors and also of chromosome 8 in mesoblastic nephroma are commonly seen. Association of specific chromosome imbalances with benign and malignant forms of papillary renal tumors, therefore, not only contribute to an understanding of tumor origins and evolution, but also implicate aneuploidy of the respective chromosomes in the tumorigenic transformation process.

In colorectal tumors, chromosome aneuploidy is a common occurrence. In fact, molecular allelotyping studies have suggested that limited karyotyping data available from these tumors actually underestimate the true extent of these changes. Losses of heterozygosity reflecting loss of the maternal or paternal allele in tumors are widespread and often accompanied by a gain of the opposite allele. Therefore, for example, a tumor could lose a maternal chromosome while duplicating the same paternal chromosome, leaving the tumor cell

with a normal karyotype and ploidy but an aberrant allelotype. It has been estimated that cancer of the colon, breast, pancreas, or prostate may lose an average of 25% of its alleles. It is not unusual to discover that a tumor has lost over half of its alleles [4]. In clinical settings, DNA ploidy measurements have revealed that DNA aneuploidy indicates high risk of developing severe premalignant changes in patients with ulcerative colitis, who are known to have an increased risk of developing colorectal cancer [11]. DNA aneuploidy has been found to be one of the useful indicators of lymph node metastasis in patients with gastric carcinoma and associated with poor outcome compared with diploid cases [12,13]. CGH analyses of chromosome aneuploidy, on the other hand, was reported to correlate gain of chromosome 20q with high tumor S phase fractions and loss of 4q with low tumor apoptotic indices [14]. Aneuploidy of chromosome 4 in metastatic colorectal cancer has recently been confirmed in studies that used unbiased DNA fingerprinting with arbitrarily primed polymerase chain reactions to detect moderate gains and losses of specific chromosomal DNA sequences [15]. The molecular karyotype (amplotype) generated from colorectal cancer revealed that moderate gains of sequences from chromosomes 8 and 13 occurred in most tumors, suggesting that overrepresentation of these chromosomal regions is a critical step for metastatic colorectal cancer.

In addition to being implicated in tumorigenesis and correlated with distinct tumor phenotypes, chromosome aneuploidy has been used as a marker of risk assessment and prognosis in several other cancers. The potential value of aneuploidy as a noninvasive tool to identify individuals at high risk of developing head and neck cancer appears especially promising. Interphase fluorescence *in situ* hybridization (FISH) revealed extensive aneuploidy in tumors from patients with head and neck squamous cell carcinomas (HNSCC) and also in clinically normal distant oral regions from the same individuals [16,17]. It has been proposed that a panel of chromosome probes in FISH analyses may serve as an important tool to detect subclinical tumorigenesis and for diagnosis of residual disease. The presence of aneuploid or tetraploid populations is seen in 90% to 95% of esophageal adenocarcinomas, and when seen in conjunction with Barrett's esophagus, a premalignant condition, predicts progression of disease [18,19]. Chromosome ploidy analyses in conjunction with loss of heterozygosity and gene mutation studies in Barrett's esophagus reflect evolution of neoplastic cell lineages *in vivo* [20]. Evolution of neoplastic progeny from Barrett's esophagus following somatic genetic mutations frequently involves bifurcations and loss of heterozygosity at several chromosomal loci leading to aneuploidy and cancer. Accordingly, it is hypothesized that during

tumor cell evolution diploid cell progenitors with somatic genetic abnormalities undergo expansion with acquired genetic instability. Such instability, often manifested in the form of increased incidence of aneuploidy, enters a phase of clonal evolution beginning in premalignant cells that proceeds over a period of time and occasionally leads to malignant transformation. The clonal evolution continues even after the emergence of cancer.

The significance of DNA and chromosome aneuploidy in other human cancers continue to be evaluated. Among papillary thyroid carcinomas, aneuploid DNA content in tumor cells was reported to correlate with distant metastases, reflecting worsened prognosis [21]. Genome wide screening of follicular thyroid tumors by CGH, on the other hand, revealed frequent loss of chromosome 22 in widely invasive follicular carcinomas [22]. Chromosome copy number gains in invasive neoplasm compared with foci of ductal carcinoma *in situ* (DCIS) with similar histology have been proposed to indicate involvement of aneuploidy in progression of human breast cancer [23]. ISH analyses of cervical intraepithelial neoplasia has provided suggestive evidence that chromosomes 1, 7 and X aneusomy is associated with progression toward cervical carcinoma [24].

Although the prognostic value of numeric aberrations remains a matter of debate in human hematopoietic neoplasia, there have been recent studies to suggest that the presence of monosomy 7 defines a distinct subgroup of acute myeloid leukemia patients [25]. It is interesting in this context that therapy-related myelodysplastic syndromes have been reported to display monosomy 5 and 7 karyotypes, reflecting poor prognosis [26].

The clinical observations, mentioned previously, are supported by *in vitro* studies in human and rodent cells in which aneuploidy is induced at early stages of transformation [27,28]. It is even suggested that aneuploidy may cause cell immortalization, in some instances, that is a critical step preceding transformation.

Finally, in an interesting study to develop transgenic mouse models of human chromosomal diseases, chromosome segment specific duplication and deletions of the genome were reported to be constructed in mouse embryonic stem cells [29]. Three duplications for a portion of mouse chromosome 11 syntenic with human chromosome 17 were established in the mouse germline. Mice with 1Mb duplication developed corneal hyperplasia and thymic tumors. The findings represent the first transgenic mouse model of aneuploidy of a defined chromosome segment that documents the direct role of chromosome aneusomy in tumorigenesis.

Aneuploidy as "dynamic cancer-causing mutation" instead of a "consequential state" in cancer

According to the hypothesis previously discussed, aneuploidy represents either a "gain of function" or "loss of function" mutation at the chromosome level with a causative influence on the tumorigenesis process. The hypothesis, however, is based only on circumstantial evidence even though existence of aneuploidy is correlated with different tumor phenotypes. The existence of numeric chromosomal alterations in a tumor does not mean that the change arose as a dynamic mutation due to genomic instability, because several factors could lead to consequential aneuploidy in tumors, also. Although aneuploidy as a dynamic mutation due to genomic instability in tumor cells would occur at a certain measurable rate per cell generation, a consequential state of aneuploidy in tumors may not occur at a predictable rate under similar conditions or in tumors with similar phenotypes. In addition to genomic instability, differences in environmental factors with selective pressure, could explain high incidence of aneuploidy and other somatic mutations in tumors compared with normal cells [4]. These include humoral, cell substratum, and cell-cell interaction differences between tumor and normal cell environments. It could be argued that despite similar rates of spontaneous aneuploidy induction in normal and tumor cells, the latter are selected to proliferate due to altered selective pressure in the tumor cell environment, whereas the normal cells are eliminated through activation of apoptosis. Alternatively, of course, one could postulate that selective expression or overexpression of anti-apoptotic proteins or inactivation of proapoptotic proteins in tumor cells may counteract default induction of apoptosis in G2/M phase cells undergoing missegregation of chromosomes. Recent demonstration of overexpression of a G2/M phase anti-apoptotic protein survivin in cancer cells [30] suggests that this protein may favor aberrant progression of aneuploid transformed cells through mitosis. This would then lead to proliferation of aneuploid cell lineages, which may undergo clonal evolution.

To ascertain that aneuploidy is a dynamic mutational event, various human tumor cell lines and transformed rodent cell lines have been analyzed for the rate of aneuploidy induction. When grown under controlled *in vitro* conditions, such conditions ensure that environmental factors do not influence selective proliferation of cells with chromosome instability. In one study, Lengauer *et al.* [31•] provided unequivocal evidence by FISH analyses that losses or gains of multiple chromosomes occurred in excess of 10^{-2} per chromosome per generation in aneuploid colorectal cancer cell lines. The study further concluded that such chromosomal instability appeared to be a dominant trait. Using another *in*

vitro model system of Chinese hamster embryo (CHE) cells, Duesberg *et al.* [32•] have also obtained similar results. With clonal cultures of CHE cells, transformed with nongenotoxic chemicals and a mitotic inhibitor, these authors demonstrated that the overwhelming majority of the transformed colonies contained more than 50% aneuploid cells, indicating that aneuploidy would have originated from the same cells that underwent transformation. All the transformed colonies tested were tumorigenic. It was further documented that the ploidy factor representing the quotient of the modal chromosome number divided by the normal diploid number, in each clone, correlated directly with the degree of chromosomal instability. Therefore, chromosomal instability was found proportional to the degree of aneuploidy in the transformed cells and the authors hypothesized that aneuploidy is a unique mechanism of simultaneously altering and destabilizing, in a massive manner, the normal cellular phenotypes. In the absence of any evidence that the transforming chemicals used in the study did not induce other somatic mutations, it is difficult to rule out the contribution of such mutations in the transformation process. These results nonetheless make a strong case for aneuploidy being a dynamic chromosome mutation event intimately associated with cancer.

Aneuploidy versus somatic gene mutation in cancer

The idea that numeric chromosome imbalance or aneuploidy is a direct cause of cancer was proposed at the turn of the century by Theodore Boveri [33]. However, the hypothesis was largely ignored over the last several decades in favor of the somatic gene mutation hypothesis, mentioned earlier. Evidence accumulating in the literature lately on specific chromosome aneusomies recognized in primary tumors, incidence of aneuploidy in cells undergoing transformation, and aneuploid tumor cells showing a high rate of chromosome instability have led to the rejuvenation of Boveri's hypothesis. The concept has recently been discussed as a "vintage wine in a new bottle" [34•]. The author points out that except for rare cancers caused by dominant retroviral oncogenes, diploidy does not seem to occur in solid tumors, whereas aneuploidy is a rule rather than exception in cancer.

Aneuploidy as an effective mutagenic mechanism driving tumor progression, on the other hand, is being recognized as a viable solution to the paradox that with known mutation rate in non-germline cells ($\sim 10^{-7}$ per gene per cell generation) tumor cell lineages cannot accumulate enough mutant genes during a human lifetime [35]. The concept is gaining significant credibility since genes that potentially affect chromosome segregation were found mutated in human cancer. Some of

these genes have also been shown to have transforming capability in *in vitro* assays. Selected recent publications describing the findings are being discussed below in reference to the mitotic targets potentially involved in inducing chromosome segregation anomalies in cells.

Potential mitotic targets and molecular mechanisms of aneuploidy

Because aneuploidy represents numeric imbalance in chromosomes, it is reasonable to expect that aneuploidy arises due to missegregation of chromosomes during cell division. There are many potential mitotic targets, which could cause unequal segregation of chromosomes (Fig. 1). Recent investigations have identified several genes involved in regulating these mitotic targets and mitotic checkpoint functions, which can be implicated in induction of aneuploidy in tumor cells. This discussion is restricted to those mitotic targets and checkpoint genes whose abnormal functioning has been observed in cancer or has been shown to cause tumorigenic transformation of cells, in recent years. The role of telomeres is discussed elsewhere in this issue. For a more detailed description of the components of mitotic machinery and their possible involvement in causing chromosome segregation abnormalities in tumor cells, readers may refer to a recently published review [36•].

Among the mitotic targets implicated in cancer, centrosome defects have been observed in a wide variety of malignant human tumors. Centrosomes play a central role in organizing the microtubule network in interphase cells and mitotic spindle during cell division. Multipolar mitotic spindles have been observed in human cancers *in situ* and abnormalities in the form of supernumerary

centrosomes, centrosomes of aberrant size and shape as well as aberrant phosphorylation of centrosome proteins have been reported in prostate, colon, brain, and breast tumors [37,38]. In view of the findings that abnormal centrosomes retain the ability to nucleate microtubules *in vitro*, it is conceivable that cells with abnormal centrosomes may missegregate chromosomes producing aneuploid cells. The molecular and genetic bases of abnormal centrosome generation and the precise pathway through which they regulate the chromosome segregation process remain to be elucidated. Recent discovery of a centrosome-associated kinase STK15/BTAK/aurora2, naturally amplified and overexpressed in human cancers, has raised the interesting possibility that aberrant expression of this kinase is critically involved in abnormal centrosome function and unequal chromosome segregation in tumor cells [39,40]. Exogenous expression of the kinase in rodent and human cells was found to correlate with an abnormal number of centrosomes, unequal partitioning of chromosomes during division, and tumorigenic transformation of cells. It is relevant in this context to mention that the *Xenopus* homologue of human STK15/BTAK/aurora2 kinase has recently been shown to phosphorylate a microtubule motor protein XIEg5, the human orthologue of which is known to participate in the centrosome separation during mitosis [41]. Findings on STK15/aurora2 kinase, thus, provide an interesting lead to a possible molecular mechanism of centrosome's role in oncogenesis. Centrosomes have, of late, been implicated in oncogenesis from studies revealing supernumerary centrosomes in *p53*-deficient fibroblasts and overexpression of another centrosome kinase PLK1 being detected in human non-small cell lung cancer [42].

One of the critical events that ensures equal partitioning of the chromosomes during mitosis is the proper and timely separation of sister chromatids that are attached to each other and to the mitotic spindle. Untimely separation of sister chromatids has been suspected as a cause of aneuploidy in human tumors. Cohesion between sister chromatids is established during replication of chromosomes and is retained until the next metaphase/anaphase transition. It has been shown that during metaphase-anaphase transition, the anaphase promoting complex/cyclosome triggers the degradation of a group of proteins called securins that inhibit sister chromatid separation. A vertebrate securin (*v*-securin) has recently been identified that inhibits sister chromatid separation and is involved in transformation and tumorigenesis. Subsequent analysis revealed that the human securin is identical to the product of the gene called pituitary tumor transforming gene, which is overexpressed in some tumors and exhibits transforming activity in NIH3T3 cells. It is proposed that elevated expression of the *v*-securin may contribute to generation of malignant tumors due to

Figure 1. Potential mitotic targets causing aneuploidy in oncogenesis

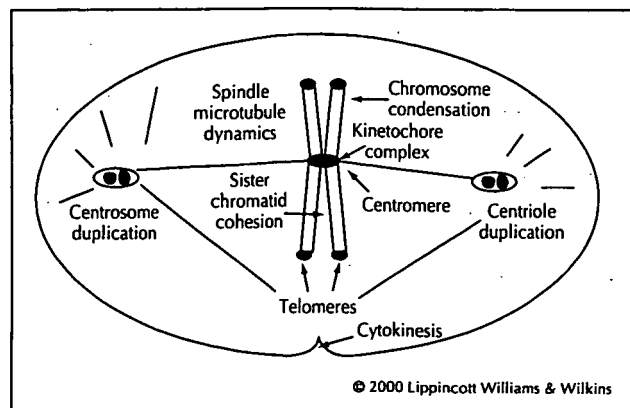


Diagram illustrates that defects in several processes involving chromosomal, spindle microtubule, and centrosomal targets, in addition to abnormal cytokinesis, may cause unequal partitioning of chromosomes during mitosis, leading to aneuploidy. Recently obtained evidence in favor of some of these possibilities is discussed in the text.

chromosome gain or loss produced by errors in chromatid separation [43•].

Normal progression through mitosis during prophase to anaphase transition is monitored at least at two checkpoints. One checkpoint operates during early prophase at G2 to metaphase progression while the second ensures proper segregation of chromosomes during metaphase to anaphase transition. Several mitotic checkpoint genes responding to mitotic spindle defects have been identified in yeast. The metaphase-anaphase transition is delayed following activation of this checkpoint during which kinetochores remain unattached to the spindle. The signal is transmitted through a kinetochore protein complex consisting of Mps1p and several Mad and Bub proteins [44]. It is expected that for unequal chromosome segregation to be perpetuated through cell proliferation cycles giving rise to aneuploidy, checkpoint controls have to be abrogated.

Following this logic, Vogelstein *et al.* [45•] hypothesized that aneuploid tumors would reveal mutation in mitotic spindle checkpoint genes. Subsequent studies by these investigators have proven the validity of this hypothesis and a small fraction of human colorectal cancers have revealed the presence of mutations in either hBub1 or hBubR1 checkpoint genes. It was further revealed that mutant BUB1 could function in a dominant negative manner conferring an abnormal spindle checkpoint when expressed exogenously. Inactivation of spindle checkpoint function in virally induced leukemia has also recently been documented following the finding that hMAD1 checkpoint protein is targeted by the Tax protein of the human T-cell leukemia virus type 1. Abrogation of hMAD1 function leads to multinucleation and aneuploidy [46].

In addition to mitotic spindle checkpoint defects, failed DNA damage checkpoint function in yeast is frequently associated with aberrant chromosome segregation as well. It, therefore, appears intriguing yet relevant that the human *BRCA1* gene, proposed to be involved in DNA damage checkpoint function, when mutated by a targeted deletion of exon 11 led to defective G2/M cell cycle checkpoint function and genetic instability in mouse embryonic fibroblasts [47]. The cells revealed multiple functional centrosomes and unequal chromosome segregation and aneuploidy. Although the molecular basis for these abnormalities is not known at this time, it raises the interesting possibility that such an aneuploidy-driven mechanism may be involved in tumorigenesis in individuals carrying germline mutations of *BRCA1* gene.

Conclusion

Growing evidence from human tumor cytogenetic investigations strongly suggest that aneuploidy is associated with the development of tumor phenotypes. Clinical findings of correlation between aneuploidy and tumorigenesis are supported by studies with *in vitro* grown transformed cell lines. Molecular genetic analyses of tumor cells provide credible evidence that mutations in genes controlling chromosome segregation during mitosis play a critical role in causing chromosome instability leading to aneuploidy in cancer. Further elucidation of molecular and physiologic bases of chromosome instability and aneuploidy induction could lead to the development of new therapeutic approaches for common forms of cancer.

Acknowledgments

The author is thankful to Drs. Bill Brinkley and Pramila Sen for discussions and advice. Help from Ms. Donna Sprabary and Ms. Hongyi Zhou in preparation of this manuscript is gratefully acknowledged. The work in the author's laboratory was supported by grants from the NIH and The University of Texas M.D. Anderson Cancer Center.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- Of special interest
- Of outstanding interest

- 1 Heim S, Mitelman F: Cancer cytogenetics, edn 2. New York: Wiley Liss Inc., 1995.
- 2 Nowell PC: The clonal evolution of tumor cell populations. *Science* 1976, 194:23-28.
- 3 Tomlinson IP, Novelli MR, Bodmer WF: The mutation rate and cancer. *Proc Natl Acad Sci USA* 1996, 93:14800-14803.
- 4 Lengauer C, Kinzler KW, Vogelstein B: Genetic instabilities in human cancers. *Nature* 1998, 396:643-649.
- An excellent review on the significance and possible mechanisms of genetic instability in cancer.
- 5 Heppner GH, Miller FR: The cellular basis of tumor progression. *Int Rev Cytol* 1998, 177:1-56.
- 6 Wolman SR: Chromosomal markers: signposts on the road to understanding neoplastic disease. *Diag Cytopath* 1998, 18:18-23.
- 7 Ross JS: DNA ploidy and cell cycle analysis in cancer diagnosis and prognosis. *Oncology* 1996, 10:867-890.
- 8 Magennis DP: Nuclear DNA in histological and cytological specimens: measurement and prognostic significance. *Br J Biomed Sci* 1997, 54:140-148.
- 9 Fletcher JA: Renal and bladder cancers. In: *Human Cytogenetic Cancer Markers*. Edited by Wolman SR, Sell S. Totowa, NJ: Humana Press; 1997:169-202.
- 10 Zhuang Z, Park WS, Pack S, Schmidt L, Vortmeyer AO, Pak E, et al.: Trisomy 7-harboring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. *Nat Genet* 1998, 20:66-69.
- 11 Lindberg JO, Stenling RB, Rutegard JN: DNA aneuploidy as a marker of premalignancy in surveillance of patients with ulcerative colitis. *Br J Surg* 1999, 86:947-950.
- 12 Sasaki O, Kido K, Nagahama S: DNA ploidy, Ki-67 and p53 as indicators of lymph node metastasis in early gastric carcinoma. *Anal Quant Cytol Histol* 1999, 21:85-88.
- 13 Abad M, Ciudad J, Rincon MR, Silva I, Paz-Bouza JI, Lopez A, et al.: DNA aneuploidy by flow cytometry is an independent prognostic factor in gastric cancer. *Anal Cell Path* 1998, 16:223-231.
- 14 DeAngelis PM, Clausen OP, Schjolberg A, Stokke T: Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their

- associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 1999, 80:526-535.
- 15 Malkhosyan S, Yasuda J, Scotto JL, Sekiya T, Yokota J, Perucho M: Molecular karyotype (amplotype) of metastatic colorectal cancer by unbiased arbitrarily primed PCR DNA fingerprinting. *Proc Natl Acad Sci (USA)* 1998, 95:10170-10175.
 - 16 Ai H, Barrera JE, Pan Z, Meyers AD, Varella-Garcia M: Identification of individuals at high risk for head and neck carcinogenesis using chromosome aneuploidy detected by fluorescence in situ hybridization. *Mut Res* 1999, 439:223-232.
 - 17 Barrera JE, Ai H, Pan Z, Meyers AD, Varella-Garcia M: Malignancy detection by molecular cytogenetics in clinically normal mucosa adjacent to head and neck tumors. *Arch Otolaryngol Head Neck Surg* 1998, 124:847-851.
 - 18 Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, et al.: 17p (p53) allelic loss, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's oesophagus. *Proc Natl Acad Sci USA* 1996, 93:7081-7084.
 - 19 Teodori L, Gohde W, Persiani M, Ferrario F, Tirindelli Danesi D, Scarpignato C, et al.: DNA/protein flow cytometry as a predictive marker of malignancy in dysplasia-free Barrett's esophagus: thirteen-year follow up study on a cohort of patients. *Cytometry* 1998, 34:257-263.
 - 20 Barrett MT, Sanchez CA, Prevo LJ, Wong DJ, Galipeau PC, Paulson TG, et al.: Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat Genet* 1999, 22:106-109.
 - 21 Sturgis CD, Caraway NP, Johnston DA, Sherman SI, Kidd L, Katz RL: Image analysis of papillary thyroid carcinoma fine needle aspirates: significant association between aneuploidy and death from disease. *Cancer* 1999, 87:155-160.
 - 22 Hemmer S, Wasenius VM, Knuutila S, Joensuu H, Franssila K: Comparison of benign and malignant follicular thyroid tumours by comparative genomic hybridization. *Br J Cancer* 1998, 78:1012-1017.
 - 23 Mendelin J, Grayson M, Wallis T, Visscher DW: Analysis of chromosome aneuploidy in breast carcinoma progression by using fluorescence in situ hybridization. *Lab Inv* 1999, 79:387-393.
 - 24 Bulten J, Poddighe PJ, Robben JC, Gemmink JH, deWilde PC, Hanselaar GAGJM: Interphase cytogenetic analysis of cervical intraepithelial neoplasia. *Am J Pathol* 1998, 152:495-503.
 - 25 Krauter J, Ganser A, Bergmann L, Raghavachar A, Hoelzer D, Lübbert M, et al.: Association between structural and numerical chromosomal aberrations in acute myeloblastic leukemia: a study by RT-PCR and FISH in 447 patients with de novo AML. *Ann Hematol* 1999, 78:265-269.
 - 26 Van Den Neste E, Louviaux I, Michaux JL, Delannoy A, Michaux L, Hagemeijer A, et al.: Myelodysplastic syndrome with monosomy 5 and/or 7 following therapy with 2-chloro-2'-deoxyadenosine. *Br J Hematol* 1999, 105:268-270.
 - 27 Namba M, Mihara K, Fushimi K: immortalization of human cells and its mechanisms. *Crit Rev Oncog* 1996, 7:19-31.
 - 28 Li R, Yerganian G, Duesberg P, Kraemer A, Willer A, Rausch C, Hehlmann R: Aneuploidy correlated 100% with chemical transformation of Chinese hamster cells. *Proc Natl Acad Sci USA* 1997, 94:14506-14511.
 - 29 Liu P, Zhang H, McLellan A, Vogel H, Bradley A: Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. *Genetics* 1998, 150:1155-1168.
 - 30 Li F, Ambrosini G, Chu EY, Plescia J, Tognini S, Marchisio PC, Altieri DC: Control of apoptosis and mitotic spindle checkpoint survivin. *Nature* 1998, 396:580-584.
 - 31 Lengauer C, Kinzler KW, Vogelstein B: Genetic instability in colorectal cancers. *Nature* 1997, 386:623-627.
Demonstrates chromosomal instability in aneuploid colorectal tumor cells.
 - 32 Duesberg P, Rausch C, Rasnick D, Hehlmann R: Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci USA* 1998, 95:13692-13697.
Correlates aneuploidy and transformation in *in vitro* grown CHE cells.
 - 33 Boveri T: Zur Frage der Entstehung maligner Tumoren. Jena, Verlag von Gustav Fischer, 1914.
 - 34 Bialy H: Aneuploidy and cancer: vintage wine in a new bottle? *Nat Biotech* 1998, 16:137-138.
Discusses the significance of aneuploidy and gene mutations in cancer.
 - 35 Orr-Weaver TL, Weinberg RA: A checkpoint on the road to cancer. *Nature* 1998, 392:223-224.
 - 36 Pihan GA, Duxsey SJ: The mitotic machinery as a source of genetic instability in cancer. *Semin Cancer Biol* 1999, 9:289-302.
Describes various components and regulatory mechanisms of mitotic machinery and possible mechanisms of chromosome missegregation in cancer.
 - 37 Pihan GA, Purohit A, Wallace J, Knecht H, Udda B, Queensberry P, Duxsey SJ: Centrosome defects and genetic instability in malignant tumors. *Cancer Res* 1998, 58:3974-3985.
 - 38 Lingle WJ, Lutz WH, Ingle JN, Maible NJ, Salisbury JL: Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. *Proc Natl Acad Sci USA* 1998, 95:2950-2955.
 - 39 Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, et al.: Tumor amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet* 1998, 20:189-193.
Describes oncogenic property of centrosome associated STK15/aurora2 kinase and its involvement in aneuploidy induction.
 - 40 Bischoff JR, Anderson L, Shu Y, Morsie K, Ng I, Chan CS, et al.: A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* 1998, 17:3052-3065.
Describes oncogenic property of STK15/aurora2 kinase and involvement in colorectal cancers.
 - 41 Giet R, Uzbekov R, Cubizolles F, Le Guellec K, Prigent C: The xenopus laevis aurora related protein kinase pEq2 associates with and phosphorylates the Kinesin related protein X1Eq5. *J Biol Chem* 1999, 274:15005-15013.
 - 42 Zimmerman W, Sparks C, Duxsey S: Amorphous no longer: the centrosome comes into focus. *Curr Opin Cell Biol* 1998, 11:122-128.
 - 43 Zou H, McGarry TJ, Bernal T, Kirschner MW: Identification of a vertebrate sister chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* 1999, 285:418-421.
Demonstrates transforming and tumorigenic function of a gene inhibiting sister chromatid separation.
 - 44 Hardwick KG: The spindle checkpoint. *Trends Genet* 1998, 14:1-4.
 - 45 Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JKV, et al.: Mutations of mitotic checkpoint genes in human cancers. *Nature* 1998, 392:300-303.
Describes mitotic checkpoint gene mutations in human colorectal cancers showing chromosome instability.
 - 46 Jin DY, Spencer F, Jeang KT: Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* 1998, 33:81-91.
 - 47 Xu X, Weaver Z, Linke SP, Li C, Gotay J, Wang XW, et al.: Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform deficient cells. *Mol Cell* 1999, 3:389-395.

Overexpression of a DEAD Box Protein (DDX1) in Neuroblastoma and Retinoblastoma Cell Lines*

(Received for publication, November 17, 1997, and in revised form, June 2, 1998)

Roseline Godbout‡, Mary Packer, and Wenjun Bie

From the Department of Oncology, Cross Cancer Institute and University of Alberta, 11560 University Ave., Edmonton, Alberta T6G1Z2, Canada

The DEAD box gene, *DDX1*, is a putative RNA helicase that is co-amplified with *MYCN* in a subset of retinoblastoma (RB) and neuroblastoma (NB) tumors and cell lines. Although gene amplification usually involves hundreds to thousands of kilobase pairs of DNA, a number of studies suggest that co-amplified genes are only overexpressed if they provide a selective advantage to the cells in which they are amplified. Here, we further characterize *DDX1* by identifying its putative transcription and translation initiation sites. We analyze *DDX1* protein levels in *MYCN/DDX1*-amplified NB and RB cell lines using polyclonal antibodies specific to *DDX1* and show that there is a good correlation with *DDX1* gene copy number, *DDX1* transcript levels, and *DDX1* protein levels in all cell lines studied. *DDX1* protein is found in both the nucleus and cytoplasm of *DDX1*-amplified lines but is localized primarily to the nucleus of nonamplified cells. Our results indicate that *DDX1* may be involved in either the formation or progression of a subset of NB and RB tumors and suggest that *DDX1* normally plays a role in the metabolism of RNAs located in the nucleus of the cell.

DEAD box proteins are a family of putative RNA helicases that are characterized by eight conserved amino acid motifs, one of which is the ATP hydrolysis motif containing the core amino acid sequence DEAD (Asp-Glu-Ala-Asp) (1–3). Over 40 members of the DEAD box family have been isolated from a variety of organisms including bacteria, yeast, insects, amphibians, mammals, and plants. The prototypic DEAD box protein is the translation initiation factor, eukaryotic initiation factor 4A, which, when combined with eukaryotic initiation factor 4B, unwinds double-stranded RNA (4). Other DEAD box proteins, such as p68, Vasa, and An3, can effectively and independently destabilize/unwind short RNA duplexes *in vitro* (5–7). Although some DEAD box proteins play general roles in cellular processes such as translation initiation (eukaryotic initiation factor 4A (4)), RNA splicing (PRP5, PRP28, and SPP81 in yeast (8–10)), and ribosomal assembly (SrmB in *Escherichia coli* (11)), the function of most DEAD box proteins remains unknown. Many of the DEAD box proteins found in higher eukaryotes are tissue- or stage-specific. For example, *PL10* mRNA is expressed only in the male germ line, and its product

has been proposed to have a specific role in translational regulation during spermatogenesis (12). Vasa and ME31B are maternal proteins that may be involved in embryogenesis (13, 14). p68, found in dividing cells (15), is believed to be required for the formation of nucleoli and may also have a function in the regulation of cell growth and division (16, 17). Other DEAD box proteins are implicated in RNA degradation, mRNA stability, and RNA editing (18–20).

The human DEAD box protein gene *DDX1*¹ was identified by differential screening of a cDNA library enriched in transcripts present in the two RB cell lines Y79 and RB522A (21). The longest *DDX1* cDNA insert isolated from this library was 2.4 kb with an open reading frame from position 1 to 2201. All eight conserved motifs characteristic of DEAD box proteins are found in the predicted amino acid sequence of *DDX1* as well as a region with homology to the heterogeneous nuclear ribonucleoprotein U, a protein believed to participate in the processing of heterogeneous nuclear RNA to mRNA (22, 23). The region of homology to heterogeneous nuclear ribonucleoprotein U spans 128 amino acids and is located between the first two conserved DEAD box protein motifs, 1a and 1b.

The proto-oncogene *MYCN* encodes a member of the MYC family of transcription factors that bind to an E box element (CACGTG) when dimerized with the MAX protein (24, 25). The *MYCN* gene is amplified and overexpressed in approximately one-third of all NB tumors (26, 27). Amplification of *MYCN* is associated with rapid tumor progression and a poor clinical prognosis (26, 27). *MYCN* overexpression is usually achieved by increasing gene copy number rather than by up-regulating basal expression of *MYCN* (27, 28). Because gene amplification involves hundreds to thousands of kilobase pairs of contiguous DNA (29–32), it is possible that co-amplification of a gene located in proximity to *MYCN* may contribute to the poor clinical prognosis of *MYCN*-amplified tumors. The *DDX1* gene maps to the same chromosomal band as *MYCN*, 2p24, and is located ~400 kb telomeric to the *MYCN* gene (33–36). All four *MYCN*-amplified RB tumor cell lines tested to date are amplified for *DDX1* (21),² while approximately two-thirds of NB cell lines and 38–68% of NB tumors are co-amplified for both genes (37–39). George *et al.* (39) found a significant decrease in the mean disease-free survival of patients with *DDX1/MYCN*-amplified NB tumors compared with *MYCN*-amplified tumors. Similarly, Squire *et al.* (38) observed a trend toward a worse clinical prognosis when both genes were amplified in the tumors of NB patients. To date, there have been no reports of a

* This work was supported by the National Cancer Institute of Canada with funds from the Canadian Cancer Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) X70649.

‡ To whom correspondence should be addressed: Dept. of Oncology, Cross Cancer Institute, 11560 University Ave., Edmonton, Alberta T6G 1Z2, Canada. Tel.: 403-432-8901; Fax: 403-432-8892.

¹ The abbreviations used are: *DDX1*, DEAD box 1; NB, neuroblastoma; RB, retinoblastoma; RACE, rapid amplification of cDNA ends; PAGE, polyacrylamide gel electrophoresis; nt, nucleotide(s); MOPS, 4-morpholinepropanesulfonic acid; bp, base pair(s); kb, kilobase(s) or kilobase pair(s).

² R. Godbout, unpublished results.

tumor amplified only for *DDX1*, and the role that this gene plays in cancer formation and progression is not known.

Because of the high rate of rearrangements in amplified DNA (31, 40), it is unlikely that a gene located ~400 kb from the *MYCN* gene will be consistently amplified as an intact unit unless its product provides a growth advantage to the cell. Based on Southern blot analysis, the *DDX1* gene extends over more than 30 kb, and there are no gross rearrangements of this gene in *DDX1*-amplified tumors (21, 38). Furthermore, there is a good correlation between *DDX1* transcript levels and gene copy number in the tumors analyzed to date. However, we need to show that *DDX1* protein is overexpressed in *DDX1*-amplified tumors if we are to entertain the possibility that this protein plays a role in the tumorigenic process. Here, we isolate and characterize the 5'-end of *DDX1* mRNA and extend the *DDX1* cDNA sequence by ~300 nt. We identify the predicted initiation codon of *DDX1* and generate antisera that specifically recognize *DDX1* protein. We analyze levels of *DDX1* protein in both *DDX1*-amplified and nonamplified RB and NB tumors and study the subcellular location of this protein in the cell.

MATERIALS AND METHODS

Library Screening—A human fetal brain cDNA library (Stratagene) was screened using a 320-bp DNA fragment from the 5'-end of the 2.4-kb *DDX1* cDNA previously described (23). Phagemids containing positive inserts were excised from λ ZAP II following the supplier's directions. The ends of the cDNA inserts were sequenced using the dideoxynucleotide chain termination method with T7 DNA polymerase (Amersham Pharmacia Biotech).

A human placenta genomic library (CLONTECH) was screened with the 5'-end of *DDX1* cDNA. Positive plaques were purified, and the genomic DNA was analyzed using restriction enzymes and Southern blotting. *EcoRI*-digested DNA fragments from these clones were subcloned into pBluescript and digested with exonuclease III and mung bean nuclease to obtain sequentially deleted clones. The exon/intron map of the 5' portion of the *DDX1* gene was obtained by comparing the sequence of *DDX1* cDNA with that of the genomic DNA.

Rapid Amplification of cDNA Ends (RACE)—We used the AmpliFINDER RACE kit (CLONTECH) to extend the 5'-end of *DDX1* cDNA. Briefly, two μ g of poly(A)⁺ RNA isolated from RB522A was reverse transcribed at 52 °C using either primer P1 or P3 (Fig. 1A). The RNA template was hydrolyzed, and excess primer was removed. A single-stranded AmpliFINDER anchor containing an *EcoRI* site was ligated to the 3'-end of the cDNA using T4 RNA ligase. The cDNA was amplified using either primer P2 or P4 (Fig. 1A) and AmpliFINDER anchor primer. RACE products were cloned into pBluescript.

Primer Extension—Poly(A)⁺ RNAs were isolated from RB and NB cell lines as described previously (21, 38). The 21-nt primers 5'-TTCGT-TCTGGGACCATGTGT-3' (primer P4 in Fig. 1A) and 5'-TGGGAC-CTAGGGCTTCTGGAC-3' (primer P3 in Fig. 1A) were end-labeled with [γ -³²P]ATP (3000 Ci/mmol; Mandel Scientific) and T4 polynucleotide kinase. Each of the labeled primers was annealed to 2 μ g of poly(A)⁺ RNA at 45 °C for 90 min, and the cDNA was extended at 42 °C for 60 min using avian myeloblastosis virus reverse transcriptase (Promega). The primer extension products were heat-denatured and run on a 8% polyacrylamide gel containing 7 M urea in 1× TBE buffer. A G + A sequencing ladder served as the size standard.

S1 Nuclease Protection Assay—The S1 nuclease protection assay to map the transcription initiation site of *DDX1* was performed as described by Favaloro *et al.* (41). The DNA probe was prepared by digesting genomic DNA spanning the upstream region of *DDX1* and exon 1 with *AvaI*, labeling the ends with [γ -³²P]ATP (3000 Ci/mmol) and polynucleotide kinase, and removing the label from one of the ends by digesting the DNA with *SphI* (Fig. 4). The RNA samples were resuspended in a hybridization mixture containing 80% formamide, 40 mM PIPES, 400 mM NaCl, 1 mM EDTA, and the heat-denatured *SphI*-*AvaI* probe labeled at the *AvaI* site. The samples were incubated at 45 °C for 16 h and digested with 3000 units/ml S1 nuclease (Boehringer Mannheim) for 60 min at 37 °C. The samples were precipitated with ethanol; resuspended in 80% formaldehyde, TBE buffer, 0.1% bromophenol blue, xylene cyanol; denatured at 90 °C for 2 min; and electrophoresed in a 7 M urea, 8% polyacrylamide gel in TBE buffer.

Northern and Southern Blot Analysis—Poly(A)⁺ RNAs were isolated from RB and NB cell lines as described previously (21, 38). Two μ g of

poly(A)⁺ RNA/lane were electrophoresed in a 6% formaldehyde, 1.5% agarose gel in MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7.0) and transferred to nitrocellulose filter in 3 M sodium chloride, 0.3 M sodium citrate. The filters were hybridized to the following DNA probes, ³²P-labeled by nick translation: (i) a 1.6-kb *EcoRI* insert from *DDX1* cDNA clone 1042 (21), (ii) a 260-bp cDNA fragment spanning the 3'-end of *DDX1* exon 1 as well as exons 2 and 3, (iii) a 160-bp fragment derived from the 5'-end of *DDX1* exon 1, and (iv) α -actin cDNA to control for lane to lane variation in RNA levels. Filters were hybridized and washed under high stringency. Southern blot analysis was as described previously (21).

Preparation of Anti-DDX1 Antiserum—To prepare antiserum to the C terminus of the *DDX1* protein, we inserted a 1.8-kb *EcoRI* fragment from bp 848 to 2668 of *DDX1* cDNA (Fig. 1B) into *EcoRI*-digested pMAL-c2 expression vector (New England Biolabs). DH5 α cells transformed with this vector were grown to mid-log phase and induced with 0.1 mM isopropyl-1-thio- β -D-thiogalactoside. The cells were harvested 3–4 h postinduction and lysed by sonication. Soluble maltose binding protein-DDX1 fusion protein was affinity-purified using amylose resin, and the maltose-binding protein was cleaved with factor Xa. The *DDX1* protein was purified on a SDS-PAGE gel, electroeluted, and concentrated. Approximately 100 μ g of protein was injected into rabbits at 4–6-week intervals. For the initial injection, the protein was dispersed in complete Freund's adjuvant (Sigma), while subsequent injections were prepared in Freund's incomplete adjuvant. Blood was collected from each rabbit 10 days after injection, and the specificity of the antiserum was tested using cell extracts from RB522A. To prepare antiserum to the N terminus of *DDX1* protein, a *DDX1* cDNA fragment from bp 268 to 851 (Fig. 1B) was inserted into pGEX-4T2 (Amersham Pharmacia Biotech). The recombinant protein produced from this construct contains the first 186 amino acids of the predicted *DDX1* sequence. Soluble glutathione S-transferase-DDX1 fusion protein was purified with glutathione-Sepharose 4B (Amersham Pharmacia Biotech). The glutathione S-transferase component of the fusion protein was cleaved with thrombin.

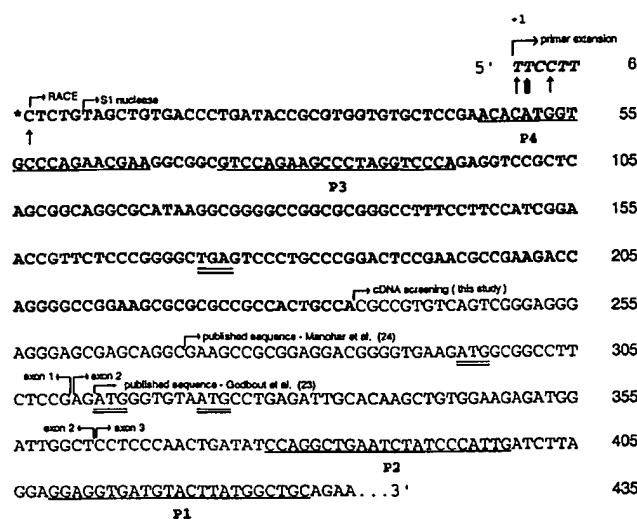
Subcellular Fractionations and Western Blot Analysis—We used two different procedures for subcellular fractionations. First, we isolated nuclear and S100 (soluble cytoplasmic) fractions from RB522A, IMR-32, Y79, RB(E)-2, HeLa, and HL60 using the procedure of Dignam (42). On average, we obtained 5–6 times more protein in the cytosolic fractions than in the nuclear fractions. Second, 10⁶ RB522A cells were lysed and fractionated into S4 (soluble cytoplasmic components), P2 (heavy mitochondria, plasma membrane fragments), P3 (mitochondria, lysosomes, peroxisomes, and Golgi membranes), and P4 fractions (membrane vesicles from rough and smooth endoplasmic reticulum, Golgi, and plasma membrane) by differential centrifugation (43). We obtained 8 mg of protein in the S4 fraction, 1 mg in P2, 0.5 mg in P3, and 2 mg in P4 fraction. The procedures related to the immunoelectron microscopy have been previously described (44).

For Western blot analysis, proteins were electrophoresed in polyacrylamide-SDS gels and electroblotted onto nitrocellulose using the standard protocol for protein transfer described by Schleicher and Schuell. The filters were incubated with a 1:5000 dilution of *DDX1* antiserum, a 1:200 dilution of anti-MYCNC monoclonal antibody (Boehringer Mannheim), or a 1:200 dilution of anti-actin (Santa Cruz Biotechnology, Inc., Santa Cruz, CA). For the colorimetric analysis, antigen-antibody interactions were visualized using either alkaline phosphatase-linked goat anti-rabbit IgG (for *DDX1*) or goat anti-mouse IgG (for MYCNC) at a 1:3000 dilution. For the ECL Western blotting analysis (Amersham Pharmacia Biotech), we used a 1:100,000 dilution of peroxidase-linked secondary anti-rabbit IgG antibody (for *DDX1*) or secondary anti-goat IgG antibody (Jackson ImmunoResearch Laboratories).

RESULTS

Identification of the 5'-End of the *DDX1* Transcript—We have previously reported the sequence of *DDX1* cDNA isolated from an RB cDNA library (21, 23). This 2.4-kb *DDX1* cDNA contains an open reading frame spanning positions 1–2201 with a methionine encoded by the first three nucleotides (Fig. 1A). There is a polyadenylation signal and poly(A) tail in the 3'-untranslated region, indicating that the sequence is complete at the 3'-end. Manohar *et al.* (37) have also isolated *DDX1* cDNA from the NB cell line LA-N-5. Their cDNA extended the 5'-end of our sequence by 42 bp and included an additional in frame methionine (*double underlined* in Fig. 1A). The possibil-

A



B



FIG. 1. Partial sequence and structure of *DDX1* cDNA. A, the sequence of the 5'-end of *DDX1* cDNA. The sequence in **boldface type** starting at the *asterisk* was obtained using the RACE strategy. The additional 6 bp in *italic boldface type* at the 5'-end of the cDNA are predicted based on the known *DDX1* genomic sequence and primer extension analysis. P1, P2, P3, and P4 are primers used in the RACE experiments (the complementary sequence was used in each case). Primers P3 and P4 were also used for the primer extension analysis. Three in frame methionine codons are indicated by the *double underline*. An in frame stop codon is indicated by the *boldface double underline*. The three major transcription initiation sites identified by primer extension are indicated by the *single arrows*, while a minor site is represented by the *broad arrow*. The predicted *DDX1* transcription initiation sites obtained by RACE, S1 nuclease, and primer extension are indicated as well as the 5'-ends of *DDX1* cDNA sequences obtained by screening cDNA libraries. The sequences transcribed from exons 1, 2, and 3 are also shown. B, the structure of the 2711-bp *DDX1* cDNA is shown with an open reading frame from position 295 to 2515.

ity of additional in frame methionines located further upstream could not be excluded, because there were no predicted stop codons in the upstream region of the cDNA.

Northern blot analysis indicated a *DDX1* transcript size of ~2800 nt, suggesting that the *DDX1* cDNAs isolated to date were lacking ~300–350 bp of 5' sequence. We have used different approaches to identify the transcription start site of *DDX1*. First, we exhaustively screened a commercial fetal brain cDNA library with the 5'-end of *DDX1* cDNA. Although numerous clones were analyzed, only one extended the sequence (by 35 bp) beyond that published by Manohar *et al.* (37) (Fig. 1A).

We next used the RACE procedure in an attempt to isolate additional 5' sequence. The nested primers used to amplify the 5'-end of the *DDX1* transcript are labeled as primers P1 and P2 in Fig. 1A and are located downstream of the three in frame methionines (*double underline* in Fig. 1A). Poly(A)⁺ RNA from RB522A was reverse transcribed at 52 °C using primer P1, and the reverse transcribed cDNA was amplified using the nested primer P2 and the 5'-RACE primer. Using this approach, we generated a product that was 230 bp longer than any of the cDNAs obtained by screening libraries (Fig. 1A). Sequencing of this 230-bp cDNA revealed an in frame stop codon (*boldface double underline* in Fig. 1A) located 123 bp

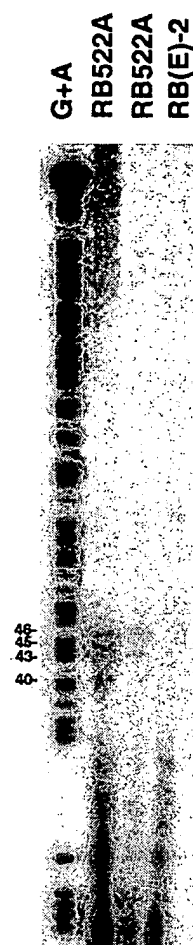


FIG. 2. Identification of the 5'-end of the *DDX1* transcript by primer extension. Radioactively labeled primer P4 was annealed to 2 µg of poly(A)⁺ RNA from RB522A (lane 1), 1 µg of poly(A)⁺ RNA from RB522A (lane 2), and 2 µg of poly(A)⁺ RNA from RB(E)-2 cells (lane 3), and extended using reverse transcriptase. The products were run on an 8% denaturing polyacrylamide gel with a G + A sequencing ladder as size marker. The primer extension products are indicated on the left. The sizes of the products (in nt) are presented as the distance from primer P4.

upstream of the predicted translation initiation site. We then prepared primers P3 and P4, located near the 5'-end of the RACE cDNA (Fig. 1A) and repeated the RACE procedure to see if additional 5' sequences could be obtained. The resulting RACE products did not extend the *DDX1* cDNA sequence further.

The location of the *DDX1* transcription initiation site was verified by primer extension. Poly(A)⁺ RNA was prepared from the following two cell lines: *DDX1*-amplified RB cell line RB522A and a nonamplified RB cell line RB(E)-2. RB522A has elevated levels of *DDX1* mRNA, while RB(E)-2 has at least 20-fold lower levels of this transcript. Three products of 40, 43, and 46 nt (with a weak signal at 45 nt) were detected in RB522A using primer P4 (Figs. 1A and 2). The 40-nt product corresponded exactly with the 5'-end of the RACE-derived cDNA while the 43- and 46-nt products extended the predicted size of the *DDX1* transcript by 3 and 6 nt, respectively. None of these products were observed in RB(E)-2. Bands of identical sizes to those obtained with RB522A mRNA were also observed in the *DDX1*-amplified NB cell line BE(2)-C but not in the *DDX1*-amplified NB cell line IMR-32 (data not shown). The same predicted *DDX1* transcription initiation site was identified with primer P3 except that the bands were of weaker



FIG. 3. Genomic map of the 5'-end of *DDX1*. The exons are represented by the black boxes, and distances are in kilobase pairs. The locations of *EcoRI* (E) sites are indicated.

intensity (data not shown). We have designated the transcription start site identified by primer extension as +1 (Fig. 1A).

The sequence of the 6 nt extending beyond the RACE cDNA was obtained by comparison of the cDNA sequence with that of *DDX1* genomic DNA. Bacteriophages containing *DDX1* genomic DNA were isolated by screening a human placenta library with 5' *DDX1* cDNA. Eighteen kb of DNA were sequenced from two bacteriophages with overlapping *DDX1* genomic DNA. Thirteen exons were identified within this 18-kb region (Fig. 3) corresponding to cDNA sequences from position 1 to 1249. The 310-bp exon 1 was by far the longest of the 13 exons sequenced, corresponding to the entire 5'-untranslated region of *DDX1* as well as the first in frame methionine. The sequences transcribed from exons 1, 2, and 3 are indicated in Fig. 1A.

Knowledge of the genomic structure of *DDX1* allowed us to use the S1 protection assay, a technique that is independent of reverse transcriptase, to further define the 5'-end of the *DDX1* transcript. Poly(A)⁺ RNAs from six *DDX1*-amplified lines (RB lines: Y79 and RB522A; NB lines: BE(2)-C, IMR-32, LA-N-1, and LA-N-5) and six nonamplified lines (RB lines: RB(E)-2 and RB412; NB lines: GOTO, NB-1, NUB-7, and SK-N-MC) were hybridized to a DNA probe that extended from position -745 in the 5'-flanking *DDX1* DNA to position +164 in exon 1. This DNA probe was labeled at position +164 as indicated in Fig. 4. Nonhybridized DNA was digested with S1 nuclease, and the sizes of the protected fragments were analyzed on a denaturing polyacrylamide gel. Bands of 150–153 nt were observed in lane 2 (RB522A), lane 5 (BE(2)-C), and lane 8 (LA-N-1) with bands of much weaker intensity in lane 7 (IMR-32) (Fig. 4). Specific bands were not detected in either *DDX1*-amplified Y79 and LA-N-5 or the nonamplified lines. Although the sizes of the S1 protected bands in RB522A, BE(2)-C, and LA-N-1 were 5 and 11 nt shorter than predicted based on RACE and primer extensions, respectively, there was general agreement with all three techniques regarding the location of the *DDX1* transcription initiation site (Fig. 1A). The smaller S1 nuclease protected products could have arisen as the result of S1 digestion of the 5'-end of the RNA:DNA heteroduplex because of its relatively high rU:dA content (45).

Identification of the same transcription initiation site in three *DDX1*-amplified lines suggests that this represents the bona fide start site of *DDX1* transcription. However, it was not clear why this start site was either very weak or not detected in three other amplified lines. To determine whether the 5'-end of exon 1 is transcribed in all *DDX1*-amplified lines, we carried out a direct analysis of the 5'-end of the *DDX1* transcript by Northern blotting. Two probes were used for this analysis: the 5' probe contained a 160-bp fragment from bp 1 to 160 (5'-half of exon 1), and the 3' probe contained a 260-bp fragment from bp 160 to 420 (3'-half of exon 1 as well as exons 2 and 3) (Fig. 1A). With the 3' probe, we obtained bands of similar size and intensity in four *DDX1*-amplified lines (RB522A, BE(2)-C, IMR-32, and LA-N-5). Band intensity was somewhat weaker in Y79 and stronger in LA-N-1 in comparison with the other lines (Fig. 5). No signal was detected in the non-*DDX1*-amplified line RB412. With the 5' probe, a relatively strong signal was observed in RB522A, BE(2)-C, and LA-N-1, while a considerably weaker

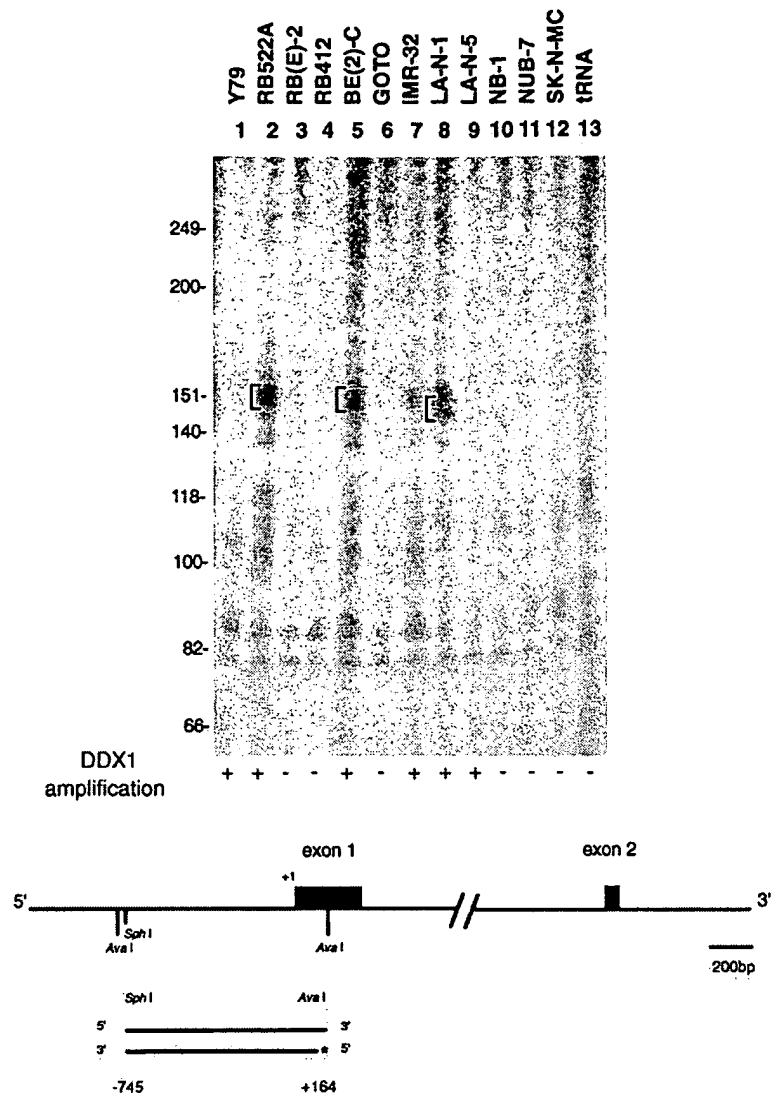
but readily apparent signal was detected in Y79, IMR-32, and LA-N-5. The signal obtained with actin indicates that, with the exception of LA-N-1, similar amounts of RNA were loaded in each lane and that the RNA was not degraded. These results indicate that at least a portion of the 160-bp 5'-end of exon 1 is transcribed in all *DDX1*-amplified lines.

Based on primer extension, S1 nuclease protection assay, Northern blot analysis and the sequencing of the RACE products, we conclude that the *DDX1* transcript is 2.7 kb with an open reading frame spanning nucleotides 295–2515 encoding a predicted protein of 740 amino acids with an estimated molecular weight of 82.4 (Fig. 1B). An in frame stop codon is located 123 nt upstream of the predicted translation initiation site, at positions 172–174. The first in frame methionine following the stop codon is in agreement with the Kozak consensus sequence (46). Furthermore, the predicted start methionine codon for human *DDX1* corresponds perfectly with that of *Drosophila* *DDX1* (47). A stop codon is located 15 nt upstream of the initiation codon in *Drosophila* *DDX1*.

Analysis of *DDX1* Protein Levels in Neuroblastoma and Retinoblastoma—We and others have previously shown that there is a good correlation between gene copy number and RNA levels in *DDX1*-amplified RB and NB cell lines (37, 38). To determine whether the correlation extends to *DDX1* protein levels, we prepared antiserum to two nonoverlapping recombinant *DDX1* proteins. First, we prepared a C terminus recombinant protein construct by inserting a 1.8-kb *EcoRI* fragment from bp 848 to 2668 (amino acids 185–740) (Fig. 1B) into the pMAL-c2 expression vector. Recombinant protein expression was induced with isopropyl-1-thio- β -D-thiogalactoside, and the 110-kDa maltose-binding protein-*DDX1* fusion product was purified by affinity chromatography using amylose resin, followed by electrophoresis on a SDS-PAGE gel after cleaving the maltose-binding protein fusion partner with factor Xa. Second, we prepared an N terminus construct by ligating a DNA fragment from bp 268 to 851 (amino acids 1–186) into pGEX-4T2. The 50-kDa glutathione S-transferase-*DDX1* fusion protein was purified by affinity chromatography on a glutathione column. This N terminus fusion protein contains only the first of the eight conserved motifs found in all DEAD box proteins, while the C terminus fusion protein includes the remaining seven motifs.

We measured *DDX1* protein levels in total cell extracts of three RB and 10 NB cell lines. Using antiserum to the N terminus fusion protein, we observed a strong signal in all *DDX1*-amplified cell lines: the RB cell lines Y79 (lane 1) and RB522A (lane 2) and the NB cell lines BE(2)-C (lane 4), IMR-32 (lane 6), LA-N-1 (lane 8), and LA-N-5 (lane 9) (Fig. 6). Two bands were observed in the majority of extracts. Of the amplified lines, Y79 produced the weakest signal, with the most intense signal observed in LA-N-1. There was an excellent correlation with *DDX1* protein and mRNA levels in these cell lines, with lower levels of *DDX1* mRNA observed in Y79 and higher levels in LA-N-1 (Fig. 7A). As shown in Fig. 7B, this correlation extended to *DDX1* gene copy number. No gross DNA rearrangements were seen in the *DDX1*-amplified lines; however, three small bands of altered size were observed in the RB412 lane. Although the nature of the DNA alteration is not known, it is noteworthy that *DDX1* transcript levels in RB412

FIG. 4. S1 nuclease mapping of the 5'-end of the *DDX1* transcript. Two μg of poly(A)⁺ RNA from four RB lines (*DDX1*-amplified Y79 and RB522A and nonamplified RB(E)-2 and RB412), eight NB lines (*DDX1*-amplified BE(2)-C, IMR-32, LA-N-1, and LA-N-5 and nonamplified GOTO, NB-1, NUB-7, and SK-N-MC), and tRNA as a negative control were hybridized to a *Sph*I–*Ava*I fragment labeled at the *Ava*I site with [γ -³²P]ATP and polynucleotide kinase. Bands of 150–153 nt are shown in lanes 2 (RB522A), 5 (BE(2)-C), and 8 (LA-N-1) with much weaker bands in lane 7 (IMR-32). A map of the probe indicating the transcription initiation site identified by primer extension (+1), the labeling site (*), and exons 1 and 2, is shown at the bottom.



are extremely low (Fig. 7A) and that the top DDX1 protein band in RB412 cell extracts is smaller in size than the top band from the other cell extracts (Fig. 6).

Two DDX1 protein bands were present in most of the lanes in Fig. 6. The same two bands were detected with antiserum to the C terminus of the DDX1 protein, as well as a third band at ~60 kDa (data not shown). There was no variation in the intensity of the 60-kDa band in *DDX1*-amplified and nonamplified cell extracts. The 60-kDa band probably represents another member of the DEAD box protein family, because the C terminus DDX1 protein used to prepare this antiserum contained seven of the eight conserved motifs found in all DEAD box proteins. To obtain an estimate of the size of the two DDX1 bands, we ran cellular extracts from RB522A on a 7% SDS-PAGE gel with the BenchMark protein ladder (Life Technologies, Inc.). The size of the DDX1 protein was determined using the Alpha Imager 2000 documentation and analysis system for molecular weight calculation. Based on this analysis, the estimated molecular mass of the top band is 89.5 kDa, while that of the bottom band is 83.5 kDa. The 84-kDa band may represent the unmodified product encoded by the *DDX1* transcript (capable of encoding a protein with a predicted molecular mass of 82.4 kDa), while the top band may represent post-translational modification of DDX1 protein (e.g. phosphorylation). Another possibility is that the top band represents intact DDX1

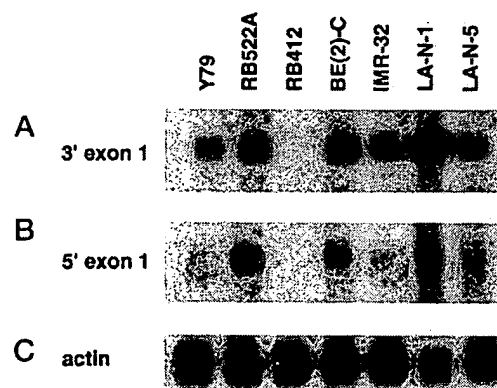


FIG. 5. Northern blot analysis of the 5'-end of the *DDX1* transcript. Two μg of poly(A)⁺ RNA isolated from *DDX1*-amplified Y79, RB522A, BE(2)-C, IMR-32, LA-N-1, and LA-N-5 and nonamplified RB412 were electrophoresed in a 1.5% agarose-formaldehyde gel. The RNA was transferred to a nitrocellulose filter and sequentially hybridized with a 260-bp fragment from *DDX1* cDNA from bp +160 to +420 (3'-end of exon 1 as well as exons 2 and 3) (A), a 160-bp fragment from *DDX1* cDNA from bp +1 to +160 (5'-end of exon 1) (B), and actin cDNA (C). The DNA was labeled with [³²P]dCTP by nick translation. The blots were hybridized and washed under high stringency.

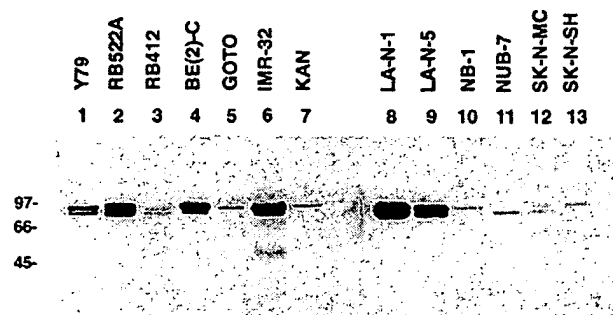


FIG. 6. DDX1 protein expression in RB and NB cell lines. Western blots were prepared using total cellular extracts from three RB (Y79, RB522A, and RB412) and 10 NB cell lines (BE(2)-C, GOTO, IMR-32, KAN, LA-N-1, LA-N-5, NB-1, NUB-7, SK-N-MC, and SK-N-SH). The lines that are amplified for the *DDX1* gene are Y79, RB522A, BE(2)-C, IMR-32, LA-N-1, and LA-N-5. Twenty μ g of protein were loaded in each lane and electrophoresed in a 10% SDS-PAGE gel. DDX1 was detected using a 1:5000 dilution of the antiserum to the amino terminus of DDX1 protein. Size markers in kilodaltons are indicated on the side.

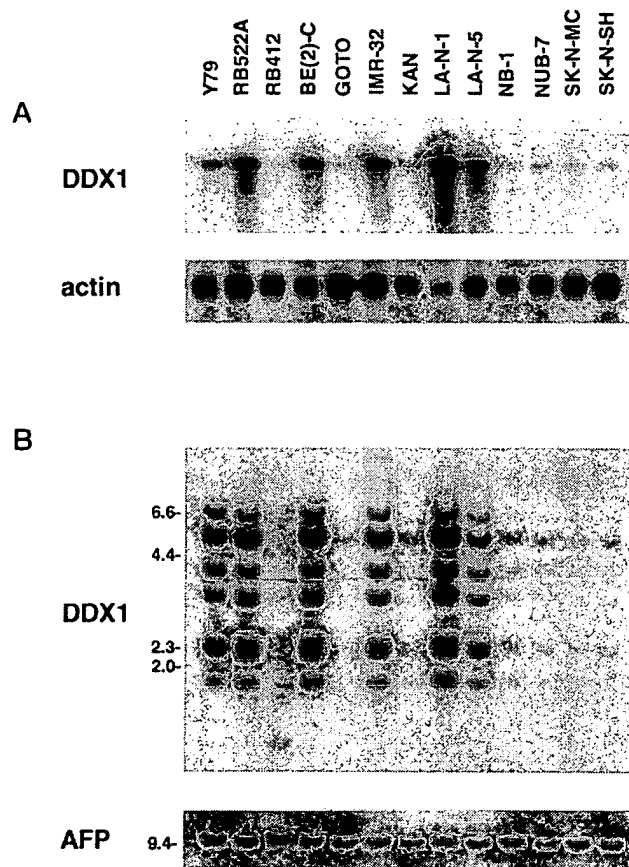


FIG. 7. Northern and Southern blot analyses of *DDX1* in RB and NB cell lines. A, 2 μ g of poly(A)⁺ RNA were loaded in each lane, electrophoresed in a 1.5% agarose-formaldehyde gel, and transferred to a nitrocellulose filter. The filter was first hybridized to a ³²P-labeled 1.6-kb *DDX1* cDNA (clone 1042) (21), stripped, and rehybridized to actin DNA. B, 10 μ g of genomic DNA from each of the indicated cell lines were digested with *Eco*RI, electrophoresed in a 1% agarose gel, and transferred to a nitrocellulose filter. The filter was hybridized to ³²P-labeled clone 1042 *DDX1* cDNA, stripped, and rehybridized with labeled α -fetoprotein cDNA. Markers (in kilobase pairs) are indicated on the side.

and the lower band is a specific truncated or degradation product of DDX1. Yet a third possibility is that the two bands represent the products of differentially spliced transcripts or

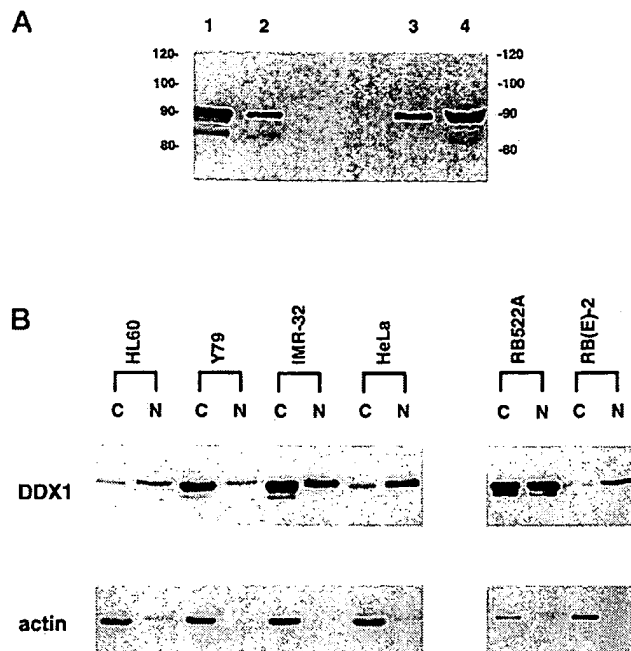


FIG. 8. Distribution of DDX1 in the nucleus and cytoplasm. A, cytosolic and nuclear extracts were prepared from RB522A and electrophoresed in a 7% SDS-PAGE gel. Cytosolic extracts were loaded in lanes 1 (20 μ g of protein) and 2 (10 μ g), while nuclear extracts were loaded in lanes 3 (10 μ g) and 4 (20 μ g). DDX1 was visualized using a 1:5000 dilution of the antiserum to the N terminus. The BenchMark protein ladder size markers (kilodaltons) are indicated on the left. B, cytosolic and nuclear extracts were prepared from HL60, Y79, IMR-32, HeLa, RB522A, and RB(E)-2 and electrophoresed in an 8% SDS-PAGE gel. Twenty μ g of proteins were loaded in each lane marked C (cytosolic) and N (nuclear). DDX1 was visualized using a 1:5000 dilution of the antiserum to the N terminus. Actin levels were analyzed using a 1:200 dilution of anti-actin antibody (Santa Cruz Biotechnology).

different translation initiation sites. However, the lack of any obvious differences in *DDX1* transcript sizes in the three RB and 10 NB lines analyzed in Fig. 7A does not support the latter possibility (e.g. compare the *DDX1* transcript size in NUB-7 (which produces the lower DDX1 protein band) and in NB-1 (which produces the higher DDX1 protein band)).

Subcellular Localization of DDX1 Protein—DEAD box proteins have been implicated in a variety of cellular functions including RNA splicing in the nucleus, translation initiation in the cytoplasm, and ribosome assembly in the nucleolus. To obtain an indication of the possible role of DDX1, we studied its subcellular location. Nuclear and cytosolic extracts were prepared from *DDX1*-amplified RB522A and run on a 7% SDS-PAGE gel. Although there was more DDX1 protein in the cytosol than in the nucleus on a per cell basis, the proportion of DDX1 protein relative to total protein was similar in both cellular compartments (Fig. 8A). Both the 90- and 84-kDa bands were present in cytosol and nuclear extracts, although the bottom band was more readily apparent in the cytosol. By running the gel for an extended period of time (twice as long as usual), we were able to detect an additional weak band at ~88 kDa in both nuclear and cytosolic extracts.

To determine whether DDX1 consistently localizes to both the cytoplasm and nucleus, we prepared cytosol and nuclear extracts from two additional *DDX1*-amplified lines, Y79 and IMR-32, as well as from nonamplified RB(E)-2, HL60, and HeLa. DDX1 protein was found in both the nucleus and cytoplasm of IMR-32, primarily in the cytoplasm of Y79, and mainly in the nucleus of the three nonamplified lines (Fig. 8B). In addition, DDX1 was almost exclusively found in nuclear

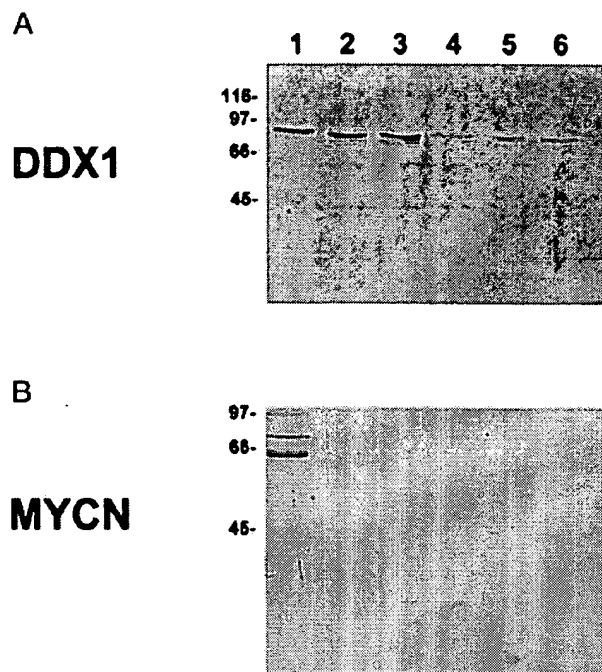


FIG. 9. Subcellular location of DDX1 protein. RB522A cells were fractionated into nuclear (lane 1), S100 and S4 cytosol (lanes 2 and 3), P2 membrane (lane 4), P3 membrane (lane 5), and P4 membrane (lane 6) fractions. Twenty μ g of protein were loaded in each lane and run on a 10% SDS-PAGE gel. A, DDX1 protein was detected using a 1:5000 dilution of the antiserum to the N terminus of DDX1. B, MYCN protein was detected using a commercially available antibody at a 1:200 dilution. Size markers (kilodaltons) are indicated on the side.

extracts prepared from normal GM38 fibroblasts (data not shown). We used anti-actin antibody to ensure that our nuclear and cytosolic extracts were not cross-contaminated (Fig. 8B).

We next carried out a more detailed analysis of DDX1 subcellular location using two different approaches: (i) fractionation of cellular components into nuclei; S100 or S4 cytosol (containing soluble cytoplasmic components, including 40 S ribosomes); P2 (heavy mitochondria, plasma membrane fragments plus material trapped by these membranes); P3 (mitochondria, lysosomes, peroxisomes, Golgi membranes, some rough endoplasmic reticulum); and P4 (microsomes from smooth and rough endoplasmic reticulum, Golgi and plasma membranes) (43); and (ii) immunogold electron microscopy. The DDX1-amplified RB522A cell line was used for both experiments. The fractionation procedures indicate that DDX1 is mainly in the nucleus and in the cytosol (S4 and S100 fractions) of RB522A cells (Fig. 9A). As a control, we used anti-human MYCN antibody to determine the location of MYCN (also amplified in RB522A) in our subcellular fractions. As shown in Fig. 9B, MYCN was primarily found in the nucleus, as one would expect of a transcription factor.

For the electron microscopy analysis, antiserum to the N terminus of DDX1 was coupled to protein A gold particles, and the distribution of DDX1 was examined in RB522A cells fixed in paraformaldehyde and glutaraldehyde. DDX1 was present in both the cytoplasm and nucleus (data not shown). There was no association with either cell organelles or with nuclear or plasma membranes.

DISCUSSION

There are presently few clues as to the function of DDX1 in normal and cancer cells. Our earlier data indicate that DDX1 mRNA is present at higher levels in fetal tissues of neural origin (retina and brain) compared with other fetal tissues (21).

There may therefore be a requirement for elevated levels of this putative RNA helicase for the efficient production or processing of neural specific transcripts. A role in cancer formation or progression is an intriguing possibility, because overexpression of an RNA unwinding protein could affect the secondary structure of RNAs in such a way as to alter the expression of specific proteins in tumor cells. DDX1 is co-amplified with MYCN in a subset of RB and NB cell lines and tumors (37–39). MYCN amplification is common in stage IV NB tumors and is a well documented indicator of poor prognosis. A general trend toward a poorer clinical prognosis is observed when both the MYCN and DDX1 genes are amplified compared with when only MYCN is amplified (38, 39), suggesting a possible role for DDX1 in NB tumor formation or progression.

It is generally accepted that co-amplified genes are not overexpressed unless they provide a selective growth advantage to the cell (48, 49). For example, although ERBA is closely linked to ERBB2 in breast cancer and both genes are commonly amplified in these tumors, ERBA is not overexpressed (48). Similarly, three genes mapping to 12q13–14 (CDK4, SAS, and MDM2) are overexpressed in a high percentage of malignant gliomas showing amplification of this chromosomal region, while other genes mapping to this region (GADD153, GLI, and A2MR) are rarely overexpressed in gene-amplified malignant gliomas (50, 51). The first three genes are probably the main targets of the amplification process, while the latter three genes are probably incidentally included in the amplicons. The data shown here indicate that DDX1 is overexpressed at both the protein and RNA levels in DDX1-amplified RB and NB cell lines and that there is a strong correlation between DDX1 gene copy number, DDX1 RNA levels, and DDX1 protein levels in these lines. Our results are therefore consistent with DDX1 overexpression playing a positive role in some aspect of NB and RB tumor formation or progression. Recently, Weiss *et al.* (52) have shown that transgenic mice that overexpress MYCN develop NB tumors several months after birth. They conclude that MYCN overexpression can contribute to the initiation of tumorigenesis but that additional events are required for tumor formation. Amplification of DDX1 may represent one of many alternative pathways by which a normal precursor “neuroblast” or “retinoblast” cell gains malignant properties.

The function of the majority of tissue-specific or developmentally regulated DEAD box genes remains unknown. However, some members of this protein family have been either directly or indirectly implicated in tumorigenesis. For example, the p68 gene has been found to be mutated in the ultraviolet light-induced murine tumor 8101 (53), while DDX6 (also known as RCK or p54) is encoded by a gene located at the breakpoint of the translocation involving chromosomes 11 and 14 in a cell line derived from a B-cell lymphoma (54, 55). Similarly, the production of a chimeric protein between DDX10 and the nucleoporin gene NUP98 has been proposed to be involved in the pathogenesis of a subset of myeloid malignancies with inv(11) (p15q22) (56). Interestingly, Grandori *et al.* (57) have shown that MYCC interacts with a DEAD box gene called MrDb, suggesting that the transcription of some DEAD box genes could be regulated through interaction with members of the MYC family. Future work will involve determining whether DDX1 represents another member of the DEAD box family with a role in the tumorigenic process.

DEAD box proteins have been implicated in translation initiation, RNA splicing, RNA degradation, and RNA stability (3, 18, 19). We carried out subcellular localization studies in an attempt to obtain a general indication of the function of DDX1. We found DDX1 protein in both the cytoplasm and nucleus of DDX1-amplified NB and RB lines. In contrast, DDX1 was

mainly located in the nucleus of nonamplified cell lines and normal fibroblast cultures. DDX1 was not associated with cellular organelles or with membranes based on immunoelectron microscopy. We therefore propose that the primary role of DDX1 is in the nucleus. The presence of DDX1 in the cytoplasm of DDX1-amplified cells may indicate that the amount of DDX1 protein that is allowed in the nucleus is tightly regulated. Alternatively, DDX1 may play a dual role in the nucleus and cytoplasm of DDX1-amplified cells.

An important component of our analysis was to identify the translation and transcription initiation sites of DDX1. We used a combination of techniques to identify the transcription start site: screening of RB and fetal brain libraries, RACE, primer extension, genomic DNA sequencing, S1 nuclease mapping, and Northern blot analysis using probes to the predicted 5'-end of the transcript. The transcription start site identified using these techniques is located ~300 nt upstream of the predicted translation initiation codon and was readily detected in three DDX1-amplified lines and barely detectable in a fourth amplified line. The 5'-untranslated region as well as the first in frame methionine are encoded within the first exon of DDX1. An in frame stop codon is located 123 nt upstream of the predicted initiation codon. We were unable to identify the transcription initiation site of DDX1 in two of the six amplified lines tested as well as in nonamplified lines. Although it remains possible that there are different transcription start sites in different cell lines, detection of lower levels (rather than the absence) of the 5'-most 160 nt of the DDX1 transcript in IMR-32, Y79, and LA-N-5 compared with RB522A, BE(2)-C, and LA-N-1 supports a quantitative rather than a qualitative difference in the 5'-end of this transcript in these cells. Our results suggest that the 5'-end of DDX1 mRNA is rarely intact, even in mRNA preparations that otherwise appear to be of high quality based on analysis of control transcripts. The 5'-end of DDX1 mRNA may therefore be especially susceptible to degradation, perhaps because of its sequence and/or secondary structure.

In conclusion, we have mapped the 5'-end of the 2.7-kb DDX1 transcript and have identified the predicted translation initiation site of DDX1 protein. We have found that DDX1-amplified RB and NB tumor lines overexpress DDX1 protein and that there is a good correlation between gene copy number and both transcript and protein levels in these cells. We have shown that DDX1 protein is primarily located in the nucleus of cells that are not DDX1-amplified. In contrast, DDX1 is present in both the nucleus and cytoplasm of DDX1-amplified NB and RB lines. A cytoplasmic location in DDX1-amplified lines may indicate that the amount of nuclear DDX1 is tightly regulated or that DDX1 plays a dual role in the cytoplasm and nucleus of these cells.

Acknowledgments—We thank Walter Dixon, Brenda Gallie, Ajay Pandita, Jeremy Squire, and Herman Yeger for the neuroblastoma and retinoblastoma cell lines. We thank Halyna Marusyk for carrying out the electron microscopy analyses. We are grateful to Randy Anderson for expert help in the preparation of the manuscript and to Stacey Hume for helpful discussions.

REFERENCES

- Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J., and Slonimski, P. P. (1989) *Nature* 337, 121–122.
- Wassarman, D. A., and Steitz, J. A. (1991) *Nature* 349, 463–464.
- Schmid, S. R., and Linder, P. (1992) *Mol. Microbiol.* 6, 283–292.
- Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C., and Sonenberg, N. (1990) *Mol. Cell. Biol.* 10, 1134–1144.
- Hirling, H., Scheffner, M., Restle, T., and Stahl, H. (1989) *Nature* 339, 562–564.
- Liang, L., Diehl-Jones, W., and Lasko, P. (1994) *Development* 120, 1201–1211.
- Gururajan, R., Mathews, L., Longo, F., and Weeks, D. L. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 2056–2060.
- Dalbadie-McFarland, G., and Abelson, J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 4236–4240.
- Jamieson, D. J., Rahe, B., Pringle, J., and Beggs, J. D. (1991) *Nature* 349, 715–717.
- Strauss, E. J., and Guthrie, C. (1994) *Nucleic Acids Res.* 22, 3187–3193.
- Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T., and Schnier, J. (1988) *Nature* 336, 496–498.
- Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D., and Fellous, M. (1989) *Cell* 57, 549–559.
- Hay, B., Jan, L. Y., and Jan, Y. N. (1988) *Cell* 55, 577–587.
- De Valoir, T., Tucker, M., Belikoff, E. J., Camp, L. A. Bolduc, C., and Beckingham, K. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 2113–2117.
- Ford, M. J., Anton, I. A., and Lane, D. P. (1988) *Nature* 332, 736–738.
- Iggo, R. D., and Lane, D. P. (1989) *EMBO J.* 8, 1827–1831.
- Buelt, M. K., Glidden, B. J., and Storm, D. R. (1994) *J. Biol. Chem.* 269, 29367–29370.
- Iost, I., and Dreyfus, M. (1994) *Nature* 372, 193–196.
- Py, B., Higgins, C. F., Krisch, H. M., and Carpousis, A. J. (1996) *Nature* 381, 169–172.
- Missel, A., Souza, A. E., Nörskau, G., and Göringer, H. U. (1997) *Mol. Cell. Biol.* 17, 4895–4903.
- Godbout, R., and Squire, J. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 7578–7582.
- Kiledjian, M., and Dreyfuss, G. (1992) *EMBO J.* 11, 2655–2664.
- Godbout, R., Hale, M., and Bisgrove, D. (1994) *Gene (Amst.)* 138, 243–245.
- Blackwood, E. M., and Eisenman, R. N. (1991) *Science* 251, 1211–1217.
- Amati, B., Dalton, S., Brooks, M. W., Littlewood, T. D., Evan, G. I., and Land, H. (1992) *Nature* 359, 423–426.
- Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E., and Bishop, J. M. (1984) *Science* 224, 1121–1124.
- Seeger, R. C., Brodeur, G. M., Sather, H., Dalton, A., Siegel, S. E., Wong, K. Y., and Hammond, D. (1985) *N. Engl. J. Med.* 313, 1111–1116.
- Cohn, S. L., Salwen, H., Quasney, M. W., Ikegaki, N., Cowan, J. M., Herst, C. V., Sharon, B., Kennett, R. H., and Rosen, S. T. (1991) *Prog. Clin. Biol. Res.* 366, 21–27.
- Cowell, J. K. (1982) *Annu. Rev. Genet.* 16, 21–59.
- Zehnbauber, B. A., Small, D., Brodeur, G. M., Seeger, R., and Vogelstein, B. (1988) *Mol. Cell. Biol.* 8, 522–530.
- Akiyama, K., and Nishi, Y. (1991) *Nucleic Acids Res.* 19, 6887–6894.
- Schneider, S. S., Hiemstra, J. L., Zehnbauber, B. A., Taillon-Miller, P., Le Paslier, D., Vogelstein, B., and Brodeur, G. M. (1992) *Mol. Cell. Biol.* 12, 5563–5570.
- Amler, L. C., Schürmann, J., and Schwab, M. (1996) *Genes Chromosomes Cancer* 15, 134–137.
- Kuroda, H., White, P. S., Sulman, E. P., Manohar, C., Reiter, J. L., Cohn, S. L., and Brodeur, G. M. (1996) *Oncogene* 13, 1561–1565.
- Noguchi, Y., Akiyama, K., Yokoyama, M., Kanda, N., Matsunaga, T., and Nishi, Y. (1996) *Genes Chromosomes Cancer* 15, 129–133.
- Pandita, A., Godbout, R., Zielenska, M., Thorner, P., Bayani, J., and Squire, J. A. (1997) *Genes Chromosomes Cancer* 20, 243–252.
- Manohar, C. F., Salwen, H. R., Brodeur, G. M., and Cohn, S. L. (1995) *Genes Chromosomes Cancer* 14, 196–203.
- Squire, J. A., Thorner, P. S., Weitzman, S., Maggi, J. D., Dirks, P., Doyle, J., Hale, M., and Godbout, R. (1995) *Oncogene* 10, 1417–1422.
- George, R. E., Kenyon, R. M., McGuckin, A. G., Malcolm, A. J., Pearson, A. D. J., and Lunec, J. (1996) *Oncogene* 12, 1583–1587.
- Reiter, J. L., and Brodeur, G. M. (1996) *Genomics* 32, 97–103.
- Favaloro, J., Treisman, R., and Kamen, R. (1980) *Methods Enzymol.* 65, 718–745.
- Dignam, J. D. (1990) *Methods Enzymol.* 182, 194–203.
- Graham, J. (1984) in *Centrifugation: A Practical Approach* (Rickwood, D. ed.) 2nd Ed., pp. 161–182, IRL Press, Oxford.
- Godbout, R., Marusyk, H., Bisgrove, D., Dabbagh, L., and Poppema, S. (1995) *Exp. Eye Res.* 60, 645–657.
- Miller, K. G., and Sollner-Webb, B. (1981) *Cell* 27, 165–174.
- Kozak, M. (1988) *Nucleic Acids Res.* 15, 8125–8132.
- Rafti, F., Scarvelis, D., and Lasko, P. F. (1996) *Gene (Amst.)* 171, 225–229.
- Van de Vijver, M., van de Bersselaar, R., Devilee, P., Cornelisse, C., Peterse, J., and Nusse, R. (1987) *Mol. Cell. Biol.* 7, 2019–2023.
- Gaudray, P., Szegetowski, P., Escot, C., Birnbaum, D., and Theillet, C. (1992) *Mutat. Res.* 276, 317–328.
- Forus, A., Florence, V. A., Maelandsmo, G. M., Meltzer, P. S., Fodstad, O., and Myklebost, O. (1993) *Cell Growth Differ.* 4, 1065–1070.
- Reifenberger, G., Reifenberger, J., Ichimura, K., Meltzer, P. S., and Collins, V. P. (1994) *Cancer Res.* 54, 4299–4303.
- Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G., and Bishop, J. M. (1997) *EMBO J.* 16, 2985–2995.
- Dubey, B. P., Hendrickson, R. C., Meredith, S. C., Siegel, C. T., Shabanowitz, J., Skipper, J. C. A., Engelhard, V. H., Hunt, D. F., and Schreiber, H. (1997) *J. Exp. Med.* 185, 695–705.
- Akao, Y., Seto, M., Yamamoto, K., Iida, S., Nakazawa, S., Inazawa, J., Abe, T., Takahashi, T., and Ueda, R. (1992) *Cancer Res.* 52, 6083–6087.
- Lu, D., and Yunis, J. J. (1992) *Nucleic Acids Res.* 20, 1967–1972.
- Arai, Y., Hosoda, F., Kobayashi, H., Arai, K., Hayashi, Y., Kamada, N., Kaneko, Y., and Ohki, M. (1997) *Blood* 89, 3936–3944.
- Grandori, C., Mac, J., Siébelt, F., Ayer, D. E., and Eisenman, R. N. (1996) *EMBO J.* 15, 4344–4357.



ORIGINAL ARTICLE

Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach

R Li¹, H Wang¹, BN Bekele², Z Yin³, NP Caraway¹, RL Katz¹, SA Stass⁴ and F Jiang⁴

¹Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ²Department of Biostatistics & Applied Mathematics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ³Department of Internal Medicine, The University of Texas Houston Medical School, Houston, TX, USA and ⁴Department of Pathology, The University of Maryland School of Medicine, Baltimore, MD, USA

Amplification and overexpression of putative oncogenes confer growth advantages for tumor development. We used a functional genomic approach that integrated simultaneous genomic and transcript microarray, proteomics, and tissue microarray analyses to directly identify putative oncogenes in lung adenocarcinoma. We first identified 183 genes with increases in both genomic copy number and transcript in six lung adenocarcinoma cell lines. Next, we used two-dimensional polyacrylamide gel electrophoresis and mass spectrometry to identify 42 proteins that were overexpressed in the cancer cells relative to normal cells. Comparing the 183 genes with the 42 proteins, we identified four genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – in which elevated protein expression correlated with both increased DNA copy number and increased transcript levels (all $r > 0.84$, two-sided $P < 0.05$). These findings were validated by Southern, Northern, and Western blotting. Specific inhibition of *EEF1A2* and *KCIP-1* expression with siRNA in the four cell lines tested suppressed proliferation and induced apoptosis. Parallel fluorescence *in situ* hybridization and immunohistochemical analyses of *EEF1A2* and *KCIP-1* in tissue microarrays from patients with lung adenocarcinoma showed that gene amplification was associated with high protein expression for both genes and that protein overexpression was related to tumor grade, disease stage, Ki-67 expression, and a shorter survival of patients. The amplification of *EEF1A2* and *KCIP-1* and the presence of overexpressed protein in tumor samples strongly suggest that these genes could be oncogenes and hence potential targets for diagnosis and therapy in lung adenocarcinoma. *Oncogene* (2006) 25, 2628–2635. doi:10.1038/sj.onc.1209289; published online 12 December 2005

Keywords: lung cancer; microarrays; proteomics; tissue microarray

Introduction

In lung adenocarcinoma, as in other types of cancer, gene amplification and the consequent overexpression of the amplified oncogene play an important role in the development of tumors, because their overexpression confers a growth advantage. The ability to identify putative oncogenes that are activated during tumorigenesis could facilitate the choice of molecular genetic targets for diagnosis and therapy of the disease. This concept has been exemplified by *HER-2*, which was first found to be amplified in neuroblastomas and subsequently shown to be associated with poor prognosis in breast cancer (Ross and Fletcher, 1999). Now, *HER-2* aberrations are used as a predictor of response to therapy, and treatment of *HER-2*-positive breast cancer with the monoclonal anti-*HER-2* antibody trastuzumab has been shown to improve prognosis (Ross and Fletcher, 1999). Emerging evidence of common amplicons in lung adenocarcinomas (Luk *et al.*, 2001; Jiang *et al.*, 2004; Tonon *et al.*, 2005) suggests that additional oncogenes remain to be identified; however, conventional techniques are ineffective in pinpointing such oncogenes. Parallel measurement of DNA copy number and mRNA levels in cDNA microarrays permits changes in copy number to be compared with transcription levels on a gene-by-gene basis to generate lists of candidate genes within the defining amplicons (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, use of transcript patterns does not allow assessment of the expression of protein products or identification of proto-oncogenes. Another approach, identifying differentially expressed proteins by proteomic analysis and then comparing the proteins present with mRNA expression in cDNA microarrays from the same specimens, can clarify the extent to which changes in transcript patterns reflect changes in their cognate proteins and post-transcriptional mechanisms (Chen *et al.*, 2002), but this approach cannot be used to identify oncogenes driven by extensive increases of their gene copy number. Moreover, using individual microarrays or proteomic approaches alone cannot distinguish the cancer-driving oncogenes that directly propel tumor progression from the larger number of passenger genes that may be concurrently over-represented but are not biologically relevant in tumor development.

Correspondence: Assistant Professor F Jiang, Department of Pathology, The University of Maryland School of Medicine, 10 South Pine Street, MSTF 7th floor, Baltimore, MD 21201-1192, USA.
E-mail: fjiang@som.umaryland.edu
Received 27 July 2005; revised 19 October 2005; accepted 27 October 2005; published online 12 December 2005

In this study, we used a comprehensive approach that integrated simultaneous comparative genomic hybridization (CGH) and transcript microarray with proteomic analyses of six lung adenocarcinoma cell lines. We directly and specifically identified four putative oncogenes that could have been activated through amplification and consequent elevation of transcript expression. We used small interfering RNA (siRNA) to inhibit the expression of two of these four genes in the lung cancer cell lines, which further implicated them in oncogenesis. We then explored the clinical significance of these findings by assessing the expression of these two genes in tissue microarrays of human lung cancer specimens. Our findings underscore the power of integrated functional genomic analyses for identifying putative oncogenes in tumorigenesis; such activated genes could be useful as targets for diagnosis or therapy in lung cancer.

Results

Simultaneous global genomic and transcript analyses identify 183 genes with increases in genomic copy numbers and transcript expression levels

To identify genes in which increased DNA copy number might contribute to increased transcript in lung adenocarcinomas, first we used CGH with microarrays of six lung adenocarcinoma cell lines. We identified 587 genes showing increases in DNA copy number across all six cell lines (Supplementary Table 1S), which were distributed as 90 amplicons on all chromosomes except for chromosomes 13 and Y (Supplementary Table 2S). A subsequent transcript test with the identical arrays of the same cell lines revealed 275 genes that showed increased mRNA levels (Supplementary Table 3S). Using random permutation tests across all cancer cell lines, we identified 183 genes (31%) that showed elevated transcript levels from the 587 genes that were over-represented in the genome (Table 1), suggesting that elevated transcript levels of the 183 genes may reflect their genomic over-representation in the cancer cells. These findings are consistent with previous reports linking genomic changes with altered transcript patterns in breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, our finding that only 31% of the genes showing increased DNA copy numbers had cognate increases in transcript expression in lung adenocarcinomas is different from the overall rates of 40–60% reported for breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). This discordance may reflect methodologic differences between studies or biological differences between breast cancer and lung adenocarcinoma.

Proteomic analyses identify four genes for which protein abundance was associated with increases in the cognate gene and transcript levels

Analysis of transcript patterns is insufficient for understanding the expression of protein products and the effect of genomic over-representation on the expression

of their cognate proteins. To extend these findings beyond genomic over-representation to expression of the protein products of those genes, we next assessed protein expression in the same cell lines by two-dimensional polyacrylamide gel electrophoresis (PAGE) and found that 42 different proteins, representing 42 individual genes, were significantly increased in the cancer cell lines (Table 2; Supplementary Figures 1S and 2S). Some of these proteins were identified as having multiple isoforms, and all individual isoforms exhibited increases in expression ranging from 4.6 to 12.8 times their expression in normal lung tissue cells. In comparing protein level of the 42 genes with changes in their cognate genomic and mRNA expression from the global microarray analyses, we found that four (9.5%) of those 42 genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – showed statistically significant correlations between elevated protein expression and increases in both copy number and mRNA expression (all $r > 0.84$; $P < 0.05$) (Table 2) in the cancer cell lines. These findings imply that the abundance of these four proteins is attributable to the amplification and consequent elevated transcription of their cognate genes.

Validation of copy number, transcript, and protein expression of PRDX1, EEF1A2, CALR, and KCIP-1 in lung cancer cell lines

To confirm our findings from the high-throughput analyses, we next used Southern, Northern, and Western blotting to assess DNA, RNA, and protein levels for the four genes identified in the six cell lines. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increase in protein level did not correlate with genetic changes. Overall, we found excellent concordance between the CGH microarray and Southern blotting analyses, transcript array and Northern blotting analyses, and proteomic and Western blotting analyses for all five genes (Figure 1). For example, *KCIP-1* showed fivefold amplification in five of the six cancer cell lines, whereas *NFKB1* showed no such increase in any of the cell lines. As for transcript expression, Northern blotting of *EEF1A2* showed high expression in five of the six cancer cell lines; again, levels of *NFKB1* transcript were not increased in any cancer cell line as compared with normal bronchial epithelial cells. The results of Western blotting were also consistent with the results of the proteomic experiments; for example, five of the cancer cell lines exhibited strong protein bands for *PRDX1* as compared with normal cells. These findings provide strong support for the validity of the results derived from the high-throughput techniques in this study.

These parallel analyses also revealed close correlations in the extent of changes in gene copies, transcript, and protein of each of the four genes in the cancer cell lines. For example, in the five cancer cell lines that showed at least fourfold increases in *EEF1A2* copy number, expression of transcript and protein was also increased by at least a factor of four as well (relative to their expression in normal cells) (Supplementary Figure 3S). The protein abundance of the four genes showing

Table 1 List of 183 genes with statistically significant correlation (0.05) between genomic copy number and transcript level

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	α
ENO1	1	8.5	0.0085
DDOST	1	20.1	0.0111
SFN	1	26.4	0.0113
MLP	1	32.2	0.0114
AKR1A1	1	45.4	0.0128
PRDX1	1	45.4	0.0122
UQCRH	1	46.2	0.0125
RPL7	1	96.4	0.0127
COL11A1	1	102.6	0.0129
MCL1	1	147.3	0.0222
PSMB4	1	148.1	0.0131
JTB	1	150.7	0.0134
RPS27	1	150.7	0.0135
HAX1	1	151	0.0266
MUC1	1	151.9	0.0143
CCT3	1	153.1	0.0167
CRABP2	1	153.4	0.0148
TKT	1	159.3	0.0152
ATP1B1	1	165.8	0.0234
CHIT1	1	199.7	0.0154
SNRPE	1	200.2	0.0165
YWHAQ	2	9.6	0.0159
ODCI	2	10.60	0.0119
RPL31	2	101.20	0.0161
BENE	2	110.40	0.0169
STAT1	2	191.80	0.0175
HSPD1	2	198.30	0.0277
HSPE1	2	198.30	0.0185
RPL37A	2	217.30	0.0388
IGFBP2	2	217.50	0.0189
RPS7	2	3.30	0.0193
RAB1A	2	65.30	0.0204
IGKC	2	89.00	0.0285
LTF	3	46.3	0.0455
PFN2	3	151	0.0207
KPNA4	3	161.5	0.0211
SI00P	4	6.7	0.1122
UGDH	4	39.3	0.0215
UCHL1	4	41.1	0.0222
SPP1	4	89.3	0.0227
TRIM2	4	154.7	0.0231
FGB	4	156	0.0235
FGG	4	156	0.0441
SDHA	5	0.251	0.0243
PDCD6	5	0.305	0.0245
CCT5	5	10.3	0.0446
PTPRF	5	14.2	0.0248
RPL37	5	40.8	0.0251
ENC1	5	74	0.0336
QP-C	5	132.2	0.0466
SPINK1	5	147.2	0.0256
CANX	5	179.2	0.0263
SOX4	6	21.7	0.0321
HDGF	6	22.6	0.0362
RPS10	6	34.6	0.0177
RPL10A	6	35.4	0.0369
VEGF	6	43.7	0.0372
OSF-2	6	45.4	0.0173
FSCN1	7	5.3	0.0378
CYCS	7	24.9	0.0381
CBX3	7	25.9	0.0289
IGFBP3	7	45.7	0.0389
CLDN4	7	72.7	0.0403
HSPB1	7	75.5	0.0433
CALR	7	92.7	0.0425
COL1A2	7	93.6	0.0457
ATPSJ2	7	98.7	0.0475
AKR1B10	7	133.6	0.0481

Table 1 (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	α
RPS20	8	56.7	0.0482
TCEB1	8	74.6	0.0486
LAPTM4B	8	98.5	0.0497
RPL30	8	98.7	0.0054
KCIP-1	8	101.6	0.0093
PABPC1	8	101.78	0.0119
EEF1D	8	144.4	0.0121
TSTA3	8	144.5	0.0122
RPL8	8	145.6	0.0128
TRA1	9	117.1	0.0136
RPL35	9	121.1	0.0133
HSPA5	9	121.5	0.0135
LCN2	9	124.4	0.0137
DPP7	9	133.4	0.0139
PFKP	10	3.2	0.0223
AKRIC1	10	5.1	0.0146
PLAU	10	75.6	0.0356
DSP	10	76.7	0.0289
TALDO1	11	0.434	0.0143
SLC22A1L	11	2.9	0.0151
TSSC3	11	2.9	0.0611
RPL27A	11	8.7	0.0156
ST5	11	8.8	0.0162
LDHA	11	18.5	0.0168
MDK	11	46.4	0.0162
DOC-1R	11	67.5	0.0167
MMP12	11	102.8	0.0177
HYOU1	11	118.9	0.0183
SCNN1A	12	6.3	0.0185
LDHB	12	21.7	0.0193
KRT7	12	52.3	0.0196
KRT5	12	52.6	0.0197
KRT6E	12	52.6	0.0201
ERBB3	12	56.2	0.0212
NACA	12	56.8	0.0218
TM4SF3	12	71.2	0.0401
NTS	12	86.2	0.0215
ASCL1	12	103.3	0.0219
TXNRD1	12	104.6	0.0223
CKAP4	12	106.6	0.0124
COX6A1	12	120.7	0.0435
BGN	12	122.5	0.0235
RAN	12	129.88	0.0238
RPL36A	14	48.1	0.0243
PGD	14	50.7	0.0248
THBS2	15	37.5	0.0251
TRAF4	15	38.3	0.0253
SPINT1	15	38.7	0.0254
RPL17	15	45.26	0.0411
PKM2	15	70.1	0.0258
IDH2	15	88.2	0.0211
RPL23A	16	0.377	0.0264
MSLN	16	0.753	0.0366
UBE2I	16	1.3	0.0271
RPS2	16	1.95	0.0281
CLDN9	16	3.1	0.0329
ARL6IP	16	18.7	0.0412
EIF3S8	16	28.3	0.0336
TUFM	16	28.9	0.0377
ALDOA	16	30.1	0.038
NME4	16	53.6	0.0381
GPR56	16	57.4	0.0386
CDH1	16	68.5	0.0289
NQO1	16	69.5	0.0396
SLC7A5	16	87.6	0.0397
APRT	16	88.6	0.0411
GALNS	16	88.6	0.0255
RPL13	16	89.3	0.0431
MCP	17	32.4	0.0465

Table 1 (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	α
ERBB2	17	35.11	0.0483
JUP	17	39.8	0.0495
CRF	17	40.39	0.0505
RPL27	17	41.1	0.0046
NME1	17	46.59	0.0082
COL1A1	17	48.6	0.0108
ABCC3	17	49.1	0.0326
NME2	17	49.6	0.0111
RPL38	17	72.7	0.0117
SMT3H2	17	73.6	0.0119
SYNGR2	17	76.6	0.0122
LGALS3BP	17	77.4	0.0127
P4HB	17	80.3	0.0126
PPAP2C	19	0.221	0.0228
GPI	19	39.55	0.0145
HPN	19	40.2	0.0129
ZNF146	19	41.4	0.0131
SPINT2	19	43.4	0.0238
PSMD8	19	43.5	0.0132
YIF1P	19	43.5	0.0135
RPS16	19	44.6	0.0144
CEACAM5	19	46.9	0.0145
CEACAM6	19	46.9	0.0143
GIPR	19	50.8	0.0259
SNRPD2	19	50.9	0.0413
KDELRL1	19	53.6	0.0152
RPL28	19	60.6	0.0156
RPS5	19	63.6	0.0267
TRIM28	19	63.7	0.0158
DAP	20	35.6	0.0166
TOP1	20	40.3	0.0172
UBE2C	20	45.1	0.0174
RPS21	20	61.6	0.0268
EEF1A2	20	62.8	0.0185
TFF3	21	42.6	0.0186
TFF1	21	42.7	0.0192
CSTB	21	44.1	0.0201
MIF	22	22.6	0.0202
XBPI	22	27.5	0.0204
PRDX4	X	22.9	0.0198
SYN1	X	46.3	0.0204
TIMP1	X	46.3	0.0209
PLP2	X	47.8	0.0212
MAGED1	X	50.3	0.0331
RPS4X	X	71	0.0124
SSR4	X	152.6	0.0232

corresponding increases in both DNA copy number and mRNA provides further evidence that these could be oncogenes, the activation of which is reflected by genomic amplification and consequent increases in transcript level in lung adenocarcinoma cell lines.

Specific inhibition of EEF1A2 and KCIP-1 expression by siRNAs led to decreased cell proliferation and induction of apoptosis

To further prove the oncogenic function of the identified genes in lung tumorigenesis, we used siRNAs to inhibit the endogenous expression of EEF1A2 and KCIP-1 protein in four lung cancer cell lines (H1563, H229, H522, and SK-LU). Transfection of the cancer cells with specific siRNAs reduced the level of EEF1A2 and KCIP-1 protein by 70–90% 48 h after transfection

(Supplementary Figure 4S). In contrast, EEF1A2 and KCIP-1 protein levels remained unchanged in mock-treated control cells and in cells transfected with a scrambled siRNA sequence. At 48 h after siRNA transfection, the percentage of proliferation of the transfected cancer cells was reduced to 15–30% as compared with 91–100% of cell proliferation of the same cell lines treated with PBS or scrambled siRNA (Supplementary Figure 5S). Apoptosis of siRNA-transfected cells was 27–34%, whereas only 4% of the same cell lines treated with PBS or scrambled siRNA showed apoptosis. These results strongly support an oncogenic role for the identified genes in lung cancer and confirm their potential usefulness as therapeutic targets for the disease.

Amplification and protein expression of KCIP-1 and EEF1A2 in lung tissue

To further validate these findings and to assess the possible clinical significance of the four potential putative oncogenes identified from the cell lines, we first applied fluorescence *in situ* hybridization and immunohistochemical analysis, in parallel, to commercially available human lung tissue microarrays (Ambion, Austin, TX, USA) to evaluate the status of two of these four genes in lung cancer tissue specimens. (Commercially available antibodies to PRDX1 or CALR were not suitable for use in immunohistochemical analysis when this report was written.) Overexpression of KCIP-1 and EEF1A2 protein in the tumors was concordant with amplification of the corresponding genes ($P=0.0003$ for KCIP-1 and $P=0.0011$ for EEF1A2). For example, 16 (35%) of the 46 lung adenocarcinomas in the microarray showed amplification of *KCIP-1*, and strong cytoplasmic staining for KCIP-1 protein was seen in 18 tumors (39%) (Figure 2). We next examined whether overexpression of these genes was associated with increased cell proliferation by analysing Ki-67 expression in contiguous sections of the tissue microarrays. Positive Ki-67 expression was found to correlate with positive expression of both KCIP-1 ($P=0.02$) and EEF1A2 ($P=0.01$). To extend these findings, we then studied 11 tissue microarray blocks comprising normal and tumor tissue specimens from 113 patients with pathologic stage I non-small-cell lung cancer who had undergone curative surgery (Wang *et al.*, 2005). Immunohistochemical analysis showed that EEF1A2 was expressed in 32 cases (28%) and KCIP-1 in 29 cases (26%). Univariate and multivariate Cox proportional hazards models were used to detect possible associations between EEF1A2 and KCIP-1 expression and clinicopathologic variables. Expression of EEF1A2 or KCIP-1 was associated with short overall survival time ($P=0.0012$ for EEF1A2 and $P=0.0026$ for KCIP-1) (Supplementary Figure 6S). Age at diagnosis, histologic type of cancer, degree of tumor differentiation, and smoking history were not associated with survival time.

Although only two genes were validated in the lung tissue microarrays (because available antibodies to the other two genes were not suitable for use in

Table 2 Proteins showing significant overexpression in cancer cell lines relative to those in normal bronchial epithelial cell lines and their correlation coefficients with increased DNA copy number or mRNA values*

Acc. no.	Gene ID	Gene	Mw/pI	Description	r with genomic copy changes ^a	r with mRNA changes ^a
Q06830	5052	PRDX1	48.4/5.4	Peroxioredoxin 1	0.92364	0.91892
Q05639	1917	EEF1A2	50.5/5.7	Eukaryotic translation elongation factor 1 alpha 2	0.90218	0.89456
P27797	811	CALR	61/5.5	Calreticulin	0.84128	0.86434
P63104	7534	KCIP-1	27/6.5	Tyrosine 3-monooxygenase activation protein, zeta	0.84467	0.85499
P07237	5034	P4HB	54/6.2	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	0.91884	0.76786
Q04695	3872	KRT17	48.0/4.9	Keratin 17	0.00236	0.86892
P09211	2950	GSTP1	23.2/4.7	Glutathione S-transferase pi	0.84218	0.69456
P17936	3486	IGFBP-3	31.6/5.8	Insulin-like growth-factor binding protein 3	0.06412	0.16434
P26641	1937	EEF1G	50/6.4	Eukaryotic translation elongation factor 1 gamma	0.00446	0.85549
P08727	3880	KRT19	44.1/5.2	Keratin 19	-0.04884	0.86786
P04792	3315	HSPB1	22/6.5	Heat shock 27 kDa protein 1	0.00364	0.31892
P00558	5230	PGK1	44.5/4.2	Phosphoglycerate kinase 1	0.50402	0.79456
Q01995	5876	TAGLN	22.5/4.3	Transgelin	-0.34128	-0.26434
P08631	3055	JTK9	59.5/6.8	Hemopoietic cell kinase	-0.01446	0.02549
P09382	3956	LGALS1	16/5.5	Galectin-1, galactoside-binding, soluble, 1	0.026623	0.01123
Q92784	8110	DPF3	25.8/4.8	D4, zinc and double PHD fingers, family 3	0.094884	-0.03214
P54257	9001	HAP1	75.5/6.5	Huntington-associated protein 1	0.12364	-0.08108
P05783	3875	KRT18	48/5.3	Keratin 18	0.010218	0.60544
P05787	3856	KRT8	9.2/4.4	Keratin 8	0.041280	0.84566
P00738	3240	HP	55.2/6.2	Haptoglobin	0.044679	-0.14501
P09769	2268	FGR	59.5/5.2	Gardner-Rasheed feline sarcoma viral oncogene homolog	0.031264	-0.13789
P19838	4790	NFKB1	50.4/6.3	Nuclear factor of kappa light gene enhancer in B-cells 1	0.04467	-0.14501
P29034	6273	S100A2	10.9/4.6	S100 calcium-binding protein A2	0.87964	0.243214
Q13105	7709	ZBTB17	87.9/5.3	Zinc-finger and BTB domain containing 17	-0.17636	0.048108
Q00987	4193	MDM2	75.2/4.8	Transformed 3T3 cell double minute 2	-0.19782	-0.50544
P27816	4134	MAP4	111/5.4	Microtubule-associated protein 4	0.25872	-0.05356
P52732	3832	KIF11	119.2/6.2	Kinesin family member 11	-0.25778	-0.53444
P25205	4172	MCM3	90.9/5.5	Minichromosome maintenance deficient 3	0.25644	0.053666
P08631	3055	HCK	59.5/5.7	Hemopoietic cell kinase	0.65533	0.054501
P09237	4316	MMP7	22.6/5.8	Matrix metalloproteinase 7	0.234987	0.876820
P30305	994	CDC25B	64.9/4.5	Cell division cycle 25B	0.045116	0.283214
P50290	998	CDC42	21.3/6.1	Cell division cycle 42 (GTP-binding protein, 25 kDa)	-0.47636	0.088108
P61586	387	RHOA	19.8/6.9	Ras homolog gene family, member A	-0.49782	-0.00544
P63000	5879	RAC1	21.5/6.8	Ras-related C3 botulinum toxin substrate 1	-0.05583	-0.03566
P07437	203068	TUBB	49.6/6.5	Tubulin, beta polypeptide	0.255533	0.145010
P24864	898	CCNE1	47.1/4.3	Cyclin E1	-0.65116	0.232149
P04141	1437	CSF2	16.9/6.3	Colony stimulating factor 2 (granulocyte-macrophage)	-0.64636	-1.28108
P28072	5694	PSMB6	25.3/5.2	Proteasome (prosome, macropain) subunit, beta type, 6	-0.69782	-1.30544
P00352	216	ALDH1A1	54.7/4.3	Aldehyde dehydrogenase 1 family, member A1	-0.75872	0.03356
Q03013	2948	GTM4	25.3/5.0	Glutathione S-transferase M4	-0.78533	0.134501
P63241	1984	EIF5A	10/4.4	Eukaryotic translation initiation factor 5A	-0.97893	-1.44321
Q01469	2171	EFABP	18.0/4.2	Fatty acid-binding protein 5	0.25684	-0.36432

*Only the gene showing statistically significant increased protein expression with increases in both genomic copy number and transcript simultaneously will be considered as potential putative oncogene in lung adenocarcinoma cells. ^ar, Spearman correlation coefficients between proteins and genomic or mRNA values are based on all six cancer cell lines; bold indicates $P < 0.05$, if $r > 0.84000$. Mw, molecular weight; pI, isoelectric point.

immunohistochemical analysis), these findings are consistent with those from our cell lines, demonstrating again that genomic amplification and consequent increases in amounts of transcript may be, at least in part, driving the abundance of proteins in these lung tumors. The association between expression of these genes and that of Ki-67, a known indicator of poor prognosis in lung cancer (Martin *et al.*, 2004), suggests that activation of these genes may be an indicator of tumor aggressiveness. These results also suggest that expression of EEF1A2 and KCIP-1 proteins in stage I non-small-cell lung cancer may be useful as a marker for distinguishing patients with relatively poor prognosis from those who might benefit from adjuvant treatment.

Discussion

Our current study illustrates the power of integrated functional genomic analyses for identifying putative oncogenes and for evaluating their potential clinical significance. Among the four identified oncogenes, three genes (*PRDX1*, *CALR*, and *KCIP-1*) have been implicated in lung tumorigenesis. *PRDX1* is an antioxidant protein involved in regulating cell proliferation, differentiation, and apoptosis. Kim *et al.* (2003) found *PRDX1* expression to be elevated in both lung cancer and adjacent normal lung tissue, suggesting that activation of *PRDX1* may enhance proliferation in lung cancer. *CALR* has a major role in Ca^{2+} binding and the

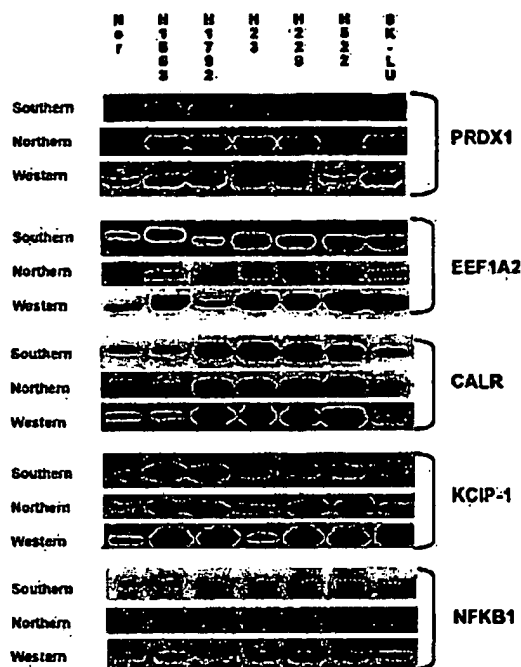


Figure 1 Confirmation by Southern, Northern, and Western blot analyses of increased DNA copies, transcript levels, and protein levels in the four genes identified in high-throughput analyses. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increased protein level did not correlate with genetic changes. The blotting results are consistent with the results from the CGH array, transcript array, and proteomic analyses. Nor, indicates normal bronchial epithelial cell line. All the experiments were repeated at least three times with each cell line. Means of normalized to β -actin signal intensities on Southern, Northern, and Western blots, along with 95% confidence intervals, were calculated (β -actin signals are not shown in the figure; two different normal bronchial epithelial cell lines were used in the confirmation and only one normal cell line is shown in the figure).

transcriptional regulation of other genes and was recently found to be overexpressed in 73% of 40 lung adenocarcinomas (Oates and Edwards, 2000). *KCIP-1* belongs to the 14-3-3 family, which participates via the MAPK and Wnt signaling pathways in the regulation of many cellular processes including cell proliferation and differentiation as well as tumorigenesis (Thomas et al., 2005). *KCIP-1* was recently found to be expressed in all 12 lung tumors tested in a single-institution study (Qi et al., 2005). Interestingly, *EEF1A2* was originally considered a putative oncogene in ovarian cancer on the basis of its being amplified in 25% and overexpressed in 30% of the same set of ovarian tumors (Anand et al., 2002); functional analyses have established its oncogenic role in cellular transformation (Lee, 2003). Our discovery that *EEF1A2* may be a putative oncogene in lung adenocarcinoma demonstrates the power of our functional genomic strategy for rapidly identifying potential oncogenes.

Although the main focus of this study was to specifically identify putative oncogenes, it should be

noted that 90.7% of the genes showing high protein expression did not show corresponding increases in both DNA copy number and transcript, a finding consistent with that of others that transcriptional, translational, and post-translational regulatory mechanisms can greatly influence the abundance of protein in lung tumorigenesis (Chen et al., 2002). For example, *NFKB1* is a critical arbiter of immune responses, cell survival, and transformation and is often activated in several types of tumors (Chen et al., 2002). Deregulation of *NFKB1* is thought to be modulated through phosphorylation of Ser337 by protein kinase A (Chen et al., 2002). In our study, 68.8% of the genes showing over-representation in the genome did not show elevated transcript levels, implying that at least some of these genes are 'passenger' genes that are concurrently amplified because of their location with respect to amplicons but lack biological relevance in terms of the development of lung adenocarcinoma.

Although the potential oncogenes we identified here are likely to be important, certainly other oncogenes could be involved in the development of lung adenocarcinoma. The oligo microarray we used consists of 22 000 probes, which represent only about 60% of the human genome. Moreover, each probe was designed for the 3' region of expressed sequence tags of the selected genes. Also, our results were initially derived from cancer cell lines, although the findings were later confirmed in human tissue samples. Our ongoing study using microarrays with information on more genes and the development of high-resolution proteomic analyses for use with larger numbers of specimens will allow more comprehensive analyses of the molecular consequences of gene amplifications. Such expanded analyses will very likely lead to the identification of additional oncogenes.

Some of the results of our current study were comparable to those of other studies of lung cancer. For example, genomic copy number and protein levels of *KCIP-1* were previously found to be amplified and overexpressed in primary lung cancers by cDNA clone-based CGH array analysis (Jiang et al., 2004) and proteomic analysis (Chen et al., 2002), respectively. Our functional genomic approach, which integrates simultaneous CGH, transcript microarrays, proteomic analyses, and siRNA, allows us not only to quickly identify potential oncogenes but also to explore their significance as diagnostic and therapeutic targets in tumor progression – more than could be achieved by any technique alone.

Genes identified in this way may serve as promising targets for diagnosis and therapy in lung adenocarcinoma. Further research on the clinical implications of such genes is needed; experiments now underway in our laboratory include overexpression of the genes in normal cells, disruption of the function of these genes in cancer cells, and investigation of how interactions among these genes (or interactions with other known oncogenes) may mediate the expression of the transformed phenotype.

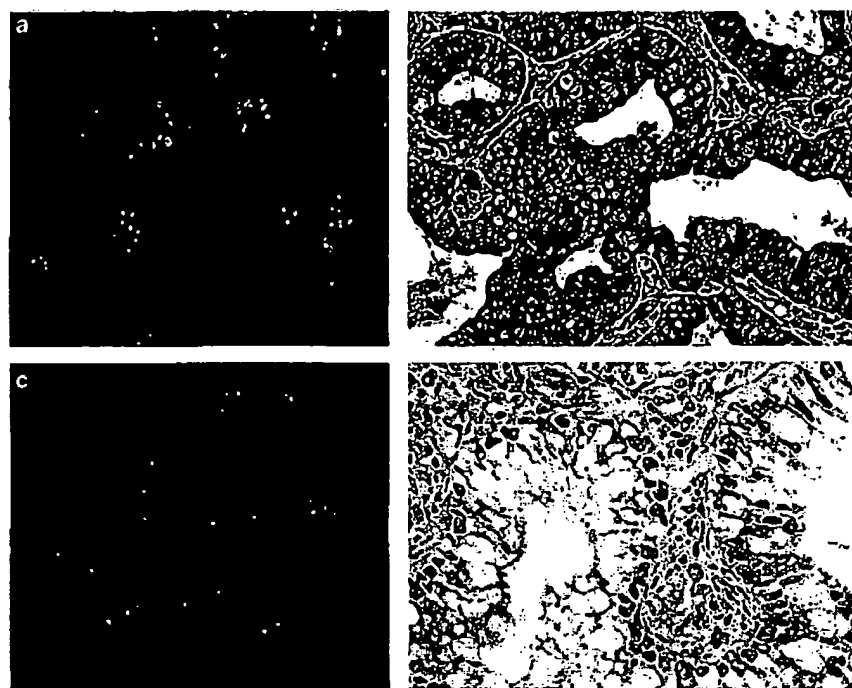


Figure 2 *EEF1A2* amplification is associated with high *EEF1A2* protein expression in lung adenocarcinomas. (a) Cells from a lung adenocarcinoma sample in which *EEF1A2* is amplified show more green signals (*EEF1A2*) than red signals (chromosome 20 centromeric probe) (original magnification, $\times 400$). (b) Immunohistochemical staining of cells from the same tissue sample as in panel a shows strong *EEF1A2* staining in the cytoplasm. (c) A lung adenocarcinoma sample with two copies of *EEF1A2* and chromosome 20 centromeric probe, indicating no *EEF1A2* amplification (original magnification, $\times 400$). (d) Immunohistochemical staining of cells from the same tissue sample as in panel c shows negative staining for *EEF1A2*.

Materials and methods

Cell lines

Six human lung adenocarcinoma cell lines (H23, H229, H1792, SK-LU-1, H522, and H1563) were obtained from the American Type Culture Collection (Manassas, VA, USA). Two normal bronchial epithelial cell lines were obtained from Clontech (Palo Alto, CA, USA). Genomic DNA, mRNA, and protein were derived from a single harvest of these cells.

DNA and RNA profiles by microarray analysis

Genomic DNA labeling and hybridization were performed as described previously (Barrett *et al.*, 2004) with Agilent's Human 1A Oligo Microarray (V2) (Agilent Technologies, Palo Alto, CA, USA), which contains 22000 unique 60-mer oligos. Details of the protocol for analysing transcripts are available at <http://www.chem.agilent.com>. Map positions for arrayed genes were assigned by identifying the DNA sequence represented in the UniGene cluster and matching it with the Golden Path genome assembly (<http://genome.ucsc.edu/>; Mat 7, 2004 Freeze). Microarray images of DNA copy number and expression were analysed by using AgilentCGH Analytics and Feature Extraction software. DNA copy number profiles that deviated significantly from background signal ratios (measured from normal control cell hybridization, as described elsewhere; Barrett *et al.*, 2004) were interpreted as evidence of true differences in DNA copy number. The criteria for defining genomic over-representation and amplicons are described elsewhere (Hyman *et al.*, 2002); details are given in the

Supplementary Information. An increase in mRNA level was defined as a twofold increase in signal ratio relative to that of the control ($\log_2 > 1$).

Quantitative two-dimensional PAGE and mass spectrometry

Analysis of proteins by two-dimensional PAGE and their identification by mass spectrometry were performed as previously described (Shen *et al.*, 2004). Briefly, protein pellets were solubilized in rehydration buffer, after which the first-dimension isoelectric focusing was carried out with a Protean IEF Cell (Bio-Rad Laboratories) and the second-dimension separation was carried out with Bio-Rad's Ready Gel Precast Gels and the Bio-Rad Criterion Cell apparatus. Protein spots were visualized by silver-based staining, and all gels were assessed with Bio-Rad's PDQuest 2D gel image analysis software. Selected spots were subjected to in-gel tryptic digestion and analysed on a Voyager-DE PRO matrix-assisted laser desorption/ionization/time-of-flight mass spectrometer (Applied Biosystems, Foster City, CA, USA). The mass list of the 20 most intense monoisotopic peaks for each sample was entered in the MS-Fit search program (v3.2.1) (<http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm>) and searched in the National Center for Biotechnology Information protein database.

Southern, Northern, and Western blot analyses

Southern, Northern, and Western blot hybridizations were performed according to standard protocols. cDNA clones for the tested genes were purchased from Invitrogen (Carlsbad,

CA, USA) and prepared as probes for the blot hybridizations. Antibodies used were obtained as follows: PRDX1, CALR, NFKB1, KCIP-1, and β -actin from Santa Cruz Biotechnology (Santa Cruz, CA, USA); and EEF1A2 from Upstate Biotechnology (Waltham, MA, USA).

Fluorescence in situ hybridization and immunohistochemical analyses of lung tissue microarrays

Fluorescence in situ hybridizations and immunohistochemical analyses of KCIP-1 and EEF1A2 were carried out as described elsewhere (Jiang et al., 2002; Wang et al., 2005) with Lung Tissue Microarrays (Ambion, Austin, TX, USA) and 11 homemade microarray blocks containing tissue samples from 113 patients with pathologic stage I non-small-cell lung cancer (Wang et al., 2005). DNA probes specific for KCIP-1 and EEF1A2 were obtained by screening a Human BAC Clone library (Invitrogen) by polymerase chain reaction as described previously (Jiang et al., 2002). The antibodies used for the immunohistochemical analyses were the same as those used for the Western blotting. Cell proliferation of the lung tissues was assessed with a Ki-67 monoclonal antibody from Santa Cruz Biotechnology. Definitions of the cutoff value for a positive result of each antibody are shown in Supplementary Information.

siRNA transfection, cellular proliferation assay, and apoptosis analysis

Transfections were carried out by using siPORT Lipid Transfection Agent (Ambion) with siRNAs targeting KCIP-1 or EEF1A2 or with a scrambled siRNA duplex (siControl) (Dharmacon Inc., Lafayette, CO, USA), with PBS used as a negative control (Jiang et al., 2002). Cells were fixed 24, 48, or 96 h later and subjected to further tests. All siRNAs were prepared by using a transcription-based method with Silencer siRNA according to the manufacturer's instructions (Ambion). Sequences of the individual siRNAs are listed in Supplementary Table 4S. Inhibition of cell growth by the

siRNAs was determined by MTT staining, and cell growth rate was plotted against the percentage of viable cells in the saline-treated controls (a value arbitrarily set at 100%) (Jiang et al., 2002). Apoptosis was analysed by fluorescence cell cycle analysis of terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling with FITC-labeled dUTP (Boehringer Mannheim Biochemicals, Mannheim, Germany) (Jiang et al., 2005).

Statistical analyses

Relationships between gene copy number and mRNA level were examined as described elsewhere (Hyman et al., 2002, Supplementary Information). Correlations between protein abundance and DNA copy number and mRNA expression of the corresponding genes were evaluated with the Spearman correlation coefficient. Fisher's exact test and χ^2 -tests were used to analyse associations between amplification and expression of the candidate genes with various histopathologic variables of the samples in the tissue microarrays. Univariate and multivariate analyses were carried out with Cox's proportional hazards model to determine which independent factors might have a joint significant influence on survival. A P -value ≤ 0.05 was considered statistically significant; all statistical tests were based on a two-sided significance level.

Acknowledgements

This work was supported by National Institutes of Health Grant CA113707-01, an institutional research grant from The University of Texas MD Anderson Cancer Center, a Developmental Project/Career Development Award from The University of Texas Specialized Programs of Research Excellence in Lung Cancer P50 CA70907, and an M Keck Center for Cancer Gene Therapy Award (all to FJ). We thank Christine F Wogan of the Department of Scientific Publications for editorial review of this manuscript.

References

- Anand N, Murthy S, Amann G, Wernick M, Porter LA, Cukier IH et al. (2002). *Nat Genet* 31: 301–305.
- Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R et al. (2004). *Proc Natl Acad Sci USA* 101: 17765–17770.
- Chen G, Gharib TG, Huang CC, Taylor JM, Misk DE, Kardia SL et al. (2002). *Mol Cell Proteomics* 1: 304–313.
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E et al. (2002). *Cancer Res* 62: 6240–6245.
- Jiang F, Caraway NP, Li RY, Katz RL. (2005). *Oncogene* 24: 3409–3418.
- Jiang F, Lin F, Price R, Gu J, Medeiros LJ, Zhang HZ et al. (2002). *J Mol Diagn* 4: 144–149.
- Jiang F, Yin Z, Caraway NP, Li R, Katz RL. (2004). *Neoplasia (New York)* 6: 623–635.
- Kim HJ, Chae HZ, Kim YJ, Kim YH, Hwang TS, Park EM et al. (2003). *Cell Biol Toxicol* 19: 285–298.
- Lee JM. (2003). *Reprod Biol Endocrinol* 1: 69–73.
- Luk C, Tsao MS, Bayani J, Shepherd F, Squire JA. (2001). *Cancer Genet Cytogenet* 125: 87–99.
- Martin B, Paesmans M, Mascaux C, Berghmans T, Lothaire P, Meert AP et al. (2004). *Br J Cancer* 91: 2018–2225.
- Oates J, Edwards C. (2000). *Histopathology* 36: 341–347.
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lønning PE et al. (2002). *Proc Natl Acad Sci USA* 99: 12963–12968.
- Qi W, Liu X, Qiao D, Martinez JD. (2005). *Int J Cancer* 113: 359–363.
- Ross JS, Fletcher JA. (1999). *Semin Cancer Biol* 9: 125–138.
- Shen J, Person MD, Zhu J, Abbuzzese JL, Li D. (2004). *Cancer Res* 64: 9018–9026.
- Thomas D, Guthridge M, Woodcock J, Lopez A. (2005). *Curr Top Dev Biol* 67: 285–303.
- Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y et al. (2005). *Proc Natl Acad Sci USA* 102: 9625–9630.
- Wang H, Zhang Z, Li R, Ang KK, Zhang H, Caraway NP et al. (2005). *Int J Cancer* 116: 285–290.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>)

Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs

Stefan Wiemann,^{1,12} Bernd Weil,^{1,2} Ruth Wellenreuther,¹
Johannes Gassenhuber,^{1,2} Sabine Glassl,³ Wilhelm Ansorge,³ Michael Böcher,⁴
Helmut Blöcker,⁴ Stefan Bauersachs,⁵ Helmut Blum,⁵ Jürgen Lauber,⁶
Andreas Düsterhöft,⁶ Andreas Beyer,⁷ Karl Köhrer,⁷ Normann Strack,²
Hans-Werner Mewes,² Birgit Ottenwälder,⁸ Brigitte Obermaier,⁸ Jens Tampe,⁹
Dagmar Heubner,¹⁰ Rolf Wambutt,¹⁰ Bernhard Korn,^{1,11} Michaela Klein,¹
and Annemarie Poustka¹

¹Molecular Genome Analysis, German Cancer Research Center, 69120 Heidelberg, Germany; ²MIPS, GSF, 82152 Martinsried, Germany; ³Biochemical Instrumentation, European Molecular Biology Laboratory, 69117 Heidelberg, Germany; ⁴GBF-Genome Analysis, 38124 Braunschweig, Germany; ⁵Genzentrum der LMU München, 81377 München, Germany; ⁶QIAGEN GmbH, 40724 Hilden, Germany; ⁷Biologisch-Medizinisches Forschungszentrum, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany; ⁸MediGenomix GmbH, 82152 Martinsried, Germany; ⁹Fraunhofer Gesellschaft, 80636 München, Germany; ¹⁰AGOWA GmbH, 12489 Berlin, Germany; ¹¹Resource Center of the German Genome Project, 69120 Heidelberg, Germany

With the complete human genomic sequence being unraveled, the focus will shift to gene identification and to the functional analysis of gene products. The generation of a set of cDNAs, both sequences and physical clones, which contains the complete and noninterrupted protein coding regions of all human genes will provide the indispensable tools for the systematic and comprehensive analysis of protein function to eventually understand the molecular basis of man. Here we report the sequencing and analysis of 500 novel human cDNAs containing the complete protein coding frame. Assignment to functional categories was possible for 52% (259) of the encoded proteins, the remaining fraction having no similarities with known proteins. By aligning the cDNA sequences with the sequences of the finished chromosomes 21 and 22 we identified a number of genes that either had been completely missed in the analysis of the genomic sequences or had been wrongly predicted. Three of these genes appear to be present in several copies. We conclude that full-length cDNA sequencing continues to be crucial also for the accurate identification of genes. The set of 500 novel cDNAs, and another 1000 full-coding cDNAs of known transcripts we have identified, adds up to cDNA representations covering 2%–5 % of all human genes. We thus substantially contribute to the generation of a gene catalog, consisting of both full-coding cDNA sequences and clones, which should be made freely available and will become an invaluable tool for detailed functional studies.

[The sequence data described in this paper have been submitted to the EMBL database under the accession nos. given in Table 2.]

The recent past has witnessed major advances in the determination of the sequence of the human genome (Dunham et al. 1999; Hattori et al. 2000). Although the whole genomic sequence will be completely unraveled in the near future (Collins et al. 1998), the identification of genes and the deciphering of gene structures will extend for a prolonged time, and cDNA sequences

will continue to be invaluable tools for this adventure, especially in view of alternative splicing. The primary focus will shift to the functional analysis of the genes and their protein products to finally understand the molecular basis of human life. Current estimates vary between 29,000 and >70,000 genes to constitute the protein coding repertoire of the human genome (Fields et al. 1994; Ewing and Green 2000; Liang et al. 2000; Roest Crolius et al. 2000). However, thus far only some 11,000 cDNA sequences have been deposited in public databases, which are supposed to contain the complete

¹²Corresponding author.

E-MAIL s.wiemann@dkfz.de; FAX 49-6221-4252-4702.

Article published on-line before print: *Genome Res.*, 10.1101/gr.154701.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.154701.

protein coding open reading frame (ORF). The majority of the respective cDNA clones are most likely not accessible. The generation of a physical clone set representing all human genes that should be made freely accessible is consequently regarded to have an extremely high impact (Schuler 1997; Pruitt et al. 2000). This would permit the establishment of a catalog of clones to provide the resources needed in the proteomics era where the functions of proteins, their action in pathways, and the possible disease relation are deciphered.

Until recently, the long-cDNA sequencing project carried out at the Kazusa Institute (Nomura et al. 1994; Nagase et al. 2000) Consortium had been the only systematic full-length cDNA sequencing project with a significant output of novel sequence information. The initiation of a new large-scale cDNA sequencing project has been announced lately that is coordinated by the National Institute of Health (Strausberg et al. 1999). We founded a cDNA Consortium in 1997 as part of the German Genome Project and aim at the characterization of the complete sequences of novel human transcripts at the cDNA level.

Here, we report the sequences and analysis of 500 novel human cDNAs that all contain the complete protein coding region. These cDNAs constitute the most valuable essence of 30,000 clones that have been EST sequenced and 3630 fully sequenced cDNAs. Over 1000 cDNAs that cover the complete coding sequence of already known transcripts have been identified in the EST-sequenced clone set. All clones are made available through the Resource Center of the German Genome Project (RZPD).

RESULTS

Libraries and Clones

To identify and sequence novel human cDNAs we have 5'-EST sequenced >30,000 independent cDNA clones. Bioinformatic evaluation of these sequences (Fig. 1) led to the identification of full-coding clones of already known proteins (>1000), and to cDNA clones lacking database hits, which are potential targets for full-length sequencing. Presumably novel cDNAs were 3'-EST sequenced and again analyzed for novelty. Out of the initial clones, 3630 cDNAs have been fully sequenced thus far, totaling 8.8 Mb. The sequence subset described here comprises 500 novel human cDNAs that are representations of the complete protein coding part of the original transcripts. Also the other fully sequenced cDNAs represent mostly genes that have not been fully sequenced elsewhere; however, the clones are not likely to contain the complete protein coding region of the respective transcripts, or they contain frame-shift mutations that have probably been introduced during reverse transcription in the cloning

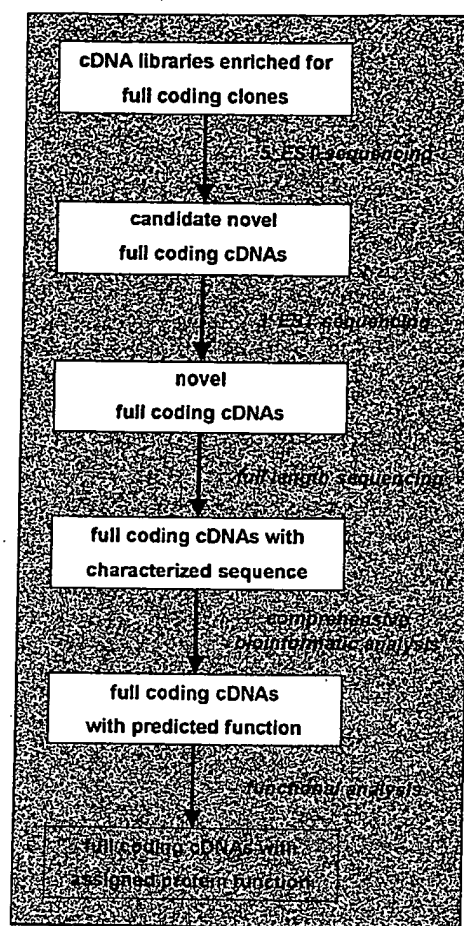


Figure 1 Flow of clones, sequences, and information in the German cDNA Consortium. 5' EST sequences were systematically generated from the clones of 384-well microtiter plates and analyzed for hits in public databases. Clones with novel sequences were 3'-EST sequenced and these ESTs were analyzed again for novelty. Clones of uncharacterized transcripts were reported back to the sequencers who then did the full-length sequencing of cDNAs. The final sequence was analyzed comprehensively with bioinformatic tools and the outputs were evaluated manually. The clones feed functional analysis projects that take advantage of the clone resources generated.

process. Therefore, these clones are only of reduced value for functional analysis. The number of bases reported for the 500 full-coding cDNAs is 1,264,620 bp; the average insert size of the clones is 2529 bp. The clones originate from five different cDNA libraries that have been sampled in varying numbers of clones (Table 1) to maximize the likelihood of identifying novel cDNAs.

The calculated average size of the encoded proteins was 470 amino acid residues, which equals the number that has been reported previously for some 1200 genes (Makalowski and Boguski 1998). There was, however, a wide variation between 66 and 1805 residues. The cDNA identifiers, the respective sequence ac-

Table 1. Library Distribution of cDNA Clones Analyzed					
RZPD library identifier	Tissue	No. of clones	% of the clones reported	Average insert size (bp)	Average ORF size (aa)
DKFZp434	Testis	204	40.8	2766	562
DKFZp564	Fetal brain	142	28.4	2049	354
DKFZp566	Fetal kidney	43	8.6	2210	328
DKFZp586	Uterus	50	10.0	2506	492
DKFZp761	Amygdala (brain)	61	12.2	3055	506

cession numbers (EMBL/GenBank/DDBJ), cDNA sizes, the length of ORFs, the chromosomal location, and functional details for the individual cDNAs are broken down in Table 2. This table is available in its entirety at <http://www.dkfz-heidelberg.de/abt0840/GCC>.

Features of 5'- and 3'-Untranslated Regions

The 5'-untranslated regions (UTRs) averaged 148 nt, which is the same range as that reported previously (Pesole et al. 1996) but considerably shorter than the number (215 nt) calculated in the UTRdb (Pesole et al. 2000). There was a wide variation in size ranging up to >800 nt (e.g., DKFZp761F182). Even this long 5'-UTR was consistent with the scanning model for translational initiation (Kozak 1999) as there was no AUG codon in this stretch of sequence. In-frame stop codons upstream from the initiator ATG were present in 56.4% (282) of the cDNAs. This number is consistent with that observed with cDNAs isolated from oligo-nucleotide cap ligation libraries (Suzuki et al. 2000), where the cDNAs have been selected to contain the extreme 5' ends of the respective transcripts. The overall GC content in the 5'-UTRs (56.3%) was considerably higher than that in the coding regions (52.6%) and the 3'-UTRs (45.7%). This is consistent with the finding that CpG islands frequently extend into the transcribed sequence (Cross and Bird 1995) whereas elements present in the 3'-UTR are often AU rich (Xu et al. 1997).

The average size of the 3'-UTRs was 926 nt [not including the poly(A) tail], which is considerably larger than the 388 nt and 820 nt reported by Makatowski and Boguski (1998) and Pesole et al. (1996), respectively. This discrepancy probably derives from the longer average size of the cDNAs described here, as compared with that observed in the previous studies. As with the 5'-UTR there was great variability with the size of the 3'-UTR. The translation terminator codon TAA could be part of the polyadenylation signal (e.g., in clone DKFZp564F2272) whereas in other cDNAs the 3'-UTR was found to be >4000 nucleotides (e.g., DKFZp486C1218).

We screened for the presence of repeat structures across the cDNA sequences. The *Alu* repeat family was

most frequently contained in the cDNAs; 7.6 % (38) of the cDNA inserts carried this type of repeat. L1 repeats were present in two cDNAs; one cDNA contained both LTR2 and *Alu* repeats (DKFZp761G18121). The repeat structures were, without exception, located in the 3'-UTR of the respective cDNAs. However, in a number of other cDNAs we found repeats also in the presumed 5'-UTRs. All of these clones turned out to be not completely spliced and/or partial upon further analysis, and having intronic sequence at the 5' ends. We therefore reason that the presence of repeat structures in 5'-UTRs of transcripts is rather rare. The lack of repeat structures in 5' EST sequences has since been implemented as criterion in the selection process of cDNA clones that are targeted to full-insert sequencing to further increase the impact of the project.

Functional Classification

We grouped the cDNAs into functional classes according to homologies of their encoded proteins with already known proteins (Table 2 and Fig. 2): cell cycle, differentiation and development, membrane protein, metabolism, nucleic acid management, protein management, signaling and communication, structure and motility, transport and traffic, and unknown. Sequence annotations in databases sometimes were misleading, and the putative function of a protein could not be simply deduced by regarding the hit with the highest similarity as being the most significant. The integration of results from several search algorithms was necessary to draw relevant conclusions. For example, the deduced protein sequences were evaluated for the presence of specific (protein) sequence patterns necessary for the function/activity of a particular protein [e.g., the DFG/DWG and aPE motifs had to be present in a protein kinase, as reported by Hanks et al. (1988)]. The results of this functional classification are given in Table 2. The largest class constitutes proteins of unknown function (202 cDNAs, 41%). Considering that for another 39 cDNAs (8%) the only prediction that had been possible was that the deduced proteins would contain a putative transmembrane domain, no function could be inferred to a total of 241 cDNAs (48%) of the predicted proteins. But even if functional

Table 2: Functional Classification of Individual cDNAs

Clone ID	cDNA data			Cell cycle	Best database hit		Tissue specificity	
	Accession no.	Contig size (bp)	ORF size (aa)		Description of best hit	Database accession no.	Gene family	Tissue specificity
DKFZP434A0530	AL136842	2768	254	2p22.1	gene: BOR2; product: GRB-containing BOR2 protein; Homo sapiens	EMBL AF164108	27e-99	
DKFZP434A1135	AL122068	3010	670	5q13	GRB-containing BOR2 protein (BOR2) mRNA; complete cds	EMBL AF076838	10	
DKFZP434A1315	AL136755	1848	387	17q21.2	Homo sapiens Rad17-like protein (RAD17) mRNA; complete cds			
					product: FIN213; the sequence of BAC FIN21 from Arabidopsis thaliana chromosome II; complete sequence	EMBL AC002130	57e-22	
DKFZP434B1741	AL501467	1546	398	15q21.1	Homo sapiens mRNA for cyclin B2; complete cds	EMBL AB020980	10	early
DKFZP434G05141	AL136750	1503	379	4p16.2	cell growth regulating nuclear protein; AR - mouse	PIRA40683	27e-144	
DKFZP434H1152	AL136840	1619	855	10p13	gene: cad2; SPB-C134710; product: cell division cycle protein 23; Homo sapiens	EMBL AL035548	7e-21	
DKFZP434J037	AL136891	3443	628	1q32.1	chromosome II (contig) B47; gene: KIAA0537; product: KIAA0537 protein; Homo sapiens	EMBL AB011109	27e-148	
					mRNA for KIAA0537 protein; complete cds			
DKFZP434N0250	AL117525	1584	462	11q13-q14	product: AKT3 protein kinase; Homo sapiens AKT3 protein kinase mRNA; complete cds	EMBL AF135794	27e-219	
DKFZP434P107	AL136894	2380	422	9q34	XPMG2 protein - African clawed frog	PIR S53818	51e-10	

Table 2. (Continued)

cDNA data			Cell cycle		Best database hit		Tissue specificity	
Clone ID	Accession no.	Contig size (bp)	ORF size (aa)	Chromosomal location	Description of best hit	Database accession no.	Gene family	Tissue score
DKFZp434P2235	AL136860	2027	549	17q12	oncogene-1 (re-2) locus (clone 210)—human gene (ORC3L) product; origin recognition complex (ORC) subunit; <i>Homo sapiens</i> origin recognition complex (ORC) subunit (ORC3L) mRNA complete cds	PIR S22155	515E-226	testis 5.181
DKFZp564A0723	AL801116	2324	712	6q14.3-q16.1	gene: <i>XA0797</i> product; mRNA for <i>XA0797</i> protein; partial cds	EMBL AF135044	0	
DKFZp564B2182	AL502611	2367	204	6q22.1-q22.33	<i>Homo sapiens</i> GGI-93 protein-mRNA complete cds	EMBL AF151856	1-2e-265	
DKFZp564B16	AL156599	1775	984	3q12.2-q12.3	gene: <i>XA0797</i> product; mRNA for <i>XA0797</i> protein; partial cds	EMBL AB010340	2.1e-50	
DKFZp564K142	AL156636	2241	335	17p11.2	<i>Rattus norvegicus</i> implantation-associated protein (AG2) mRNA; partial cds	EMBL AF008554	9.4e-184	
DKFZp564L0562	AL800090	941	185	4q31.1-21	<i>Homo sapiens</i> mRNA for APC10 complete cds	EMBL AB012109	4.4e-178	
DKFZp564N0182	AL502645	1616	144	3p21.1	<i>Homo sapiens</i> DRRI1 (DRRI) mRNA complete cds	EMBL AF099853	0	brain 6.116
DKFZp564N0182	AL502645	1646	144	3p21.1	<i>Homo sapiens</i> DRRI1 (DRRI) mRNA complete cds	EMBL AF099853	0	retina 5.45
DKFZp566G0346	AL156710	4503	262	9q22.1	<i>Homo sapiens</i> spindlein mRNA complete cds	EMBL AF106682	0	

The cDNAs have been grouped into ten functional categories (see Statistics—Classification) based on sequence similarity data and have been grouped accordingly. The cDNA clones are available from the Resource Center of the German Genome Project using the clone ID shown in the first column. The respective sequences are available at the EMBL/GenBank/DBJ databases under the accession numbers shown in the second column. The third column provides the size of the individual cDNA inserts, and the fourth column shows the size of the encoded/predicted proteins. The chromosomal location of the respective genes is shown in the fifth column. Columns 6–8 describe database hits with the highest similarity. The accession number of the best hit (and the database where this hit was found) is the description of the best hit, and the P value of this hit is provided in these three columns, respectively. Similarities were predicted based on BLASTX and BLASTN analyses. Selection of the representative "best hit" was done using the following criteria: (i) BLASTX hit was judged better than BLASTN hit; (2) in cases where the best BLASTX (only with EMBL database hit) had been calculated from the same nucleotide sequence, the hit was the best hit in the BLASTN analysis; the BLASTN hit is given; and (3) only when no other hits were available, genome sequence entries are given. The classification of a protein to a major gene family was possible (based on similarity information), the respective family is shown in column 9. Based on the availability of EST sequences from that particular tissue (at the time of analysis), respectively, this fraction is depicted from the full table (available on-line at <http://www.dkfz-heidelberg.de/abi0840/GCC>).

The cDNAs have been grouped into ten functional categories (see Statistics—Classification) based on sequence similarity data and have been grouped accordingly. The cDNA clones are available from the Resource Center of the German Genome project using the clone IDs shown in the first column. The respective sequences are available at the EMBL/GenBank/DBJ databases under the accession numbers shown in the second column. The third column provides the size of the individual cDNA inserts, and the fourth column shows the size of the encoded/predicted proteins. The chromosomal location for the respective genes is shown in the fifth column. Columns 6–8 describe database hits with the highest similarity. The accession number of the best hit, and the description of the best hit, and the P -value of this hit is provided in these three columns, respectively. Similarities were predicted based on BLASTX- and BLASTN-analyses. Selection of the representative best hit was done using the following criteria: (1) BLASTX hit was judged better than a BLASTN hit; (2) in cases where the best BLASTX (only with EMBL database) hit had been calculated from the same nucleotide sequence identity, that was the best hit in the BLASTN analysis; the BLASTN hit is given; and (3) Only when no other hits were available, genomic sequence entries are given. Classification of a protein to a major gene family was possible (based on similarity information), the respective family is shown in column 9. Based on the availability of EST information, tissue specific expression of transcripts has been depicted in columns 10–13 showing the tissue (an arbitrary score) (see WWW2001b) and the absolute number of ESTs sequenced from that particular tissue (at the time of analysis), respectively. This section is excerpted from the full table available on the web: <http://www.gdz.helldorfer.de/ab08.01/GCC>.

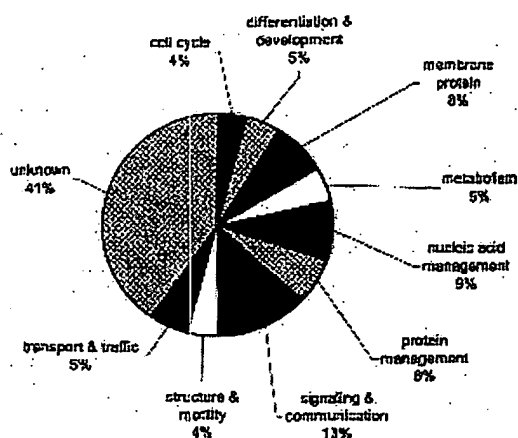


Figure 2 Functional classification of proteins encoded by the cDNAs. The deduced proteins were grouped into 10 functional categories based on sequence similarity with proteins of known function. The fraction of the 500 cDNAs grouped into the respective categories is indicated.

predictions were possible, the identification, for example, of a protein kinase, neither provides information on its substrates nor on the pathway(s) in which it is involved. Comprehensive functional analyses should be specifically indicated for a set of cDNAs encoding candidates for genes related to disease, such as putative GTP binding proteins, ion channels, and a cDNA encoding a protein that is highly similar to an oncogene.

We further analyzed the cDNAs for the presence of function-related sequence motifs to also identify novel members of gene families. We identified 41 potential leucine zipper proteins (Struhl 1989), 11 proteins with WD-domains (Neer et al. 1994), 11 proteins with predicted zinc finger domains (Parraga et al. 1988), 7 potential protein kinases, and 5 RNA helicases. The respective clones are indicated in Table 2 (column 9). Two cDNAs (DKFZp586I021 and DKFZp434O1826) contain both a WD-domain and a leucine zipper. A zinc-finger domain is predicted additionally for the deduced protein of the former cDNA.

Alternative Splicing

We found 39 (7.8%) cDNAs to represent putative splice variants of already known transcripts. This number is likely to represent the lower end of the fraction of transcripts that are alternatively spliced in vivo as any cDNAs representing already fully-known transcripts were excluded from further sequencing and alternative splice forms should therefore be under-represented in our set. We found ORFs with additional exons (e.g., DKFZp761B192), skipped exons (e.g., DKFZp564A032), and alternative exons including one containing the translational start codon and resulting in a different N terminus of the deduced peptide (e.g., DKFZp434J154).

The percentage of alternatively spliced cDNAs appeared to be slightly higher in fetal brain, 40% of the alternatively spliced cDNAs originate from fetal brain whereas only 28% of all cDNAs analyzed originate from this tissue. This finding is consistent with reports by Sutcliffe and Milner (1988) and Hanke et al. (1999). The presence of intron sequences reminiscent in many cDNA sequences available in public databases, however, might lead to an overestimation of the extent of alternative splicing that is taking place in vivo. Experimental evidence will therefore be needed to confirm presumed alternative splice forms.

Representation of cDNAs in the UniGene Data Set

Depending on the true number of human genes, about 60%–90% have already been identified by partial sequencing of >2,000,000 cDNAs (EST sequencing). Overlapping EST sequences have been clustered to break down this large number of ESTs to comprehensive collections that should consist of nonredundant data sets having one representation (cluster) for every gene. The most widely accepted clustering data set is the UniGene (Schuler et al. 1996) resource at the NCBI (<http://www.ncbi.nlm.nih.gov/UniGene/>). This dataset currently consists of >90,000 clusters of mostly partial sequences. Consensus sequences of these clusters are available from <http://www.rzpd.de>. To investigate the representation of the novel cDNAs reported here in the UniGene data set and to evaluate the maximum number of genes that could be represented there, we aligned the full-length sequences with the UniGene database. The version of UniGene (Build 105) that was used in the analysis consisted of 92,931 clusters with 10,501 clusters containing known genes.

In total, 626 UniGene clusters matched with 472 out of the 500 full-coding cDNA sequences. The majority of cDNAs (342, 68%) was represented by one UniGene cluster. An additional 130 (26%) cDNAs were represented by 284 separate UniGene clusters (Fig. 3). Thus, a number of UniGene clusters could be linked by the full-length cDNA sequences. An example of three UniGene clusters that were joined with one cDNA is given in Figure 4. We analyzed the ESTs and clusters that were placed internal to the cDNAs reported here and found that most of the EST clones making up these clusters had originated from internal priming events (mostly in reminiscent intron sequences) and not from alternative polyadenylation. The number of 640 clusters that was hit with 472 cDNA sequences implies that there is ~35% redundancy in UniGene. As the average size of the human transcripts in general has been estimated to be in the same range as the average size of the cDNAs reported here (by quantification of Northern blots that had been hybridized with a labeled oligonucleotide dT probe; N. Nomura, pers. comm.), our finding should be representative. However, the true

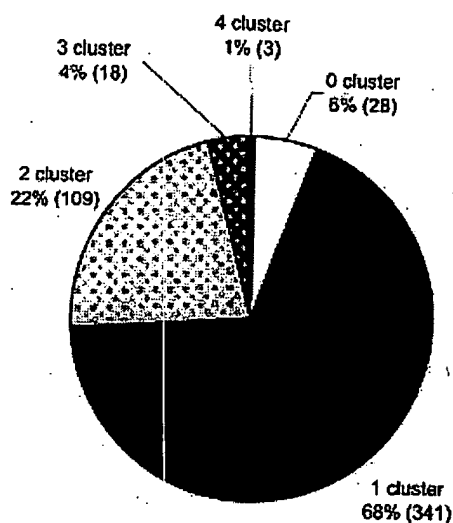


Figure 3 Representation of cDNAs in the UniGene data set (Build 105). Every cDNA was aligned with the UniGene data set to identify the number of EST clusters that was hit/joined with a given cDNA. The fraction and the total number (in parentheses) of the cDNAs are given for the varying numbers of clusters being hit.

number of genes represented in UniGene will further condense as a considerable fraction of the UniGene clusters are singletons (~39%), which are clusters made up by only one cDNA, and several of these will eventually turn out to be artifacts. Consequently, we estimate the number of independent genes that are represented in UniGene to be 50,000 at most.

A fraction of 6% (28 cDNAs) did not have hits in the UniGene database (cutoff, sequence identity >95% in 50 bp). The low number of the novel cDNAs without UniGene matches might in turn imply that >90% of all human genes were already represented in this database. However, we would rather assume that an unknown number of genes has escaped cloning and/or identification so far as the respective transcripts might be expressed only at extremely low levels or in very specialized cell types or differentiation stages. A proper selection of tissues or even single cell types for cDNA

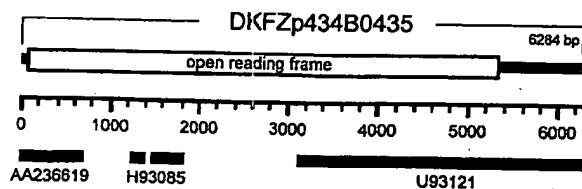


Figure 4 Three UniGene clusters are joined when aligned with the cDNA sequence DKFZp434B0435. The bar on top of the scale represents the cDNA with the open reading frame drawn as an open box. The bars below the scale represent the position and size (in bp) of the three UniGene clusters that are joined by the cDNA sequence. The accession nos. of representative sequences of the respective UniGene clusters are given below the bars.

library production will be a critical issue for the detection and cloning also of these rarely expressed transcripts. For example, fetal brain, although very complex in expression, has been so deeply sampled in EST projects [especially the IMAGE 1NIB library (Soares et al. 1994)] but also in full-length cDNA sequencing (Nagase et al. 2000) that the novelty rate (3 of 142 cDNAs, 2%) is rather low in this tissue. In contrast, testis currently appears to have a higher potential for identifying transcripts not yet covered by ESTs (19 of 204 cDNAs, 9%).

Tissue Specificity of Expression

To analyze for a possible tissue specificity of expression we aligned the cDNA sequences with the EST database dbEST. ESTs originating from pooled tissues and tissues with unclear origin were excluded. Each cDNA received a score indicating the degree of tissue specificity. The higher this score, the higher the likelihood that expression of the particular transcript should be restricted to that tissue. A ubiquitously expressed transcript would have had a score of one. Only cDNAs with scores of five or higher are indicated in Table 2 (columns 10–12). In total, the expression of 22 transcripts appeared to be restricted to only one tissue with matching tissues of our cDNA and the ESTs (Table 2). Six brain-derived cDNAs only matched ESTs that had derived from brain tissues. Most of the cDNAs encode proteins that are either involved in the cell cycle or signaling pathways, for example, a stathmin-like protein and a protein similar to a calmodulin-binding protein. Only one of the six cDNAs encodes a protein of unknown function. Another 15 testis cDNAs had hits only with ESTs from testis/male genital tract. Although predictions could be made for three of the encoded proteins (a predicted sperm flagellar protein, a putative neurotransmitter transporter, and a possible nuclear pore protein), the other 12 cDNAs encode proteins of unknown function. The only uterus cDNA predicted to be specifically expressed in uterus/ovary encodes a putative chaperone-associated protease, which could indicate that this protein might be involved in the differentiation of the egg or embryo. The expression of several testis-derived transcripts appeared to be very selective as the scores calculated for these cDNAs were rather high, compared with scores obtained with other cDNAs and tissues (Table 2). This also matches the observation that the novelty rate, counting cDNAs without EST hits, was highest in the testis library (see above).

cDNAs Mapping to Human Chromosomes 21 and 22

To demonstrate the power of mapping genes by aligning cDNA with genomic sequences we downloaded the sequences of the first two completely sequenced human chromosomes 21 (Hattori et al. 2000) and 22

(Dunham et al. 1999) and aligned them with those novel cDNAs mapping to the respective chromosomes (Table 3). Clone identifiers of the respective cDNAs and the insert and ORF sizes are provided in the first three columns. For ORF sizes (column 3) the predicted number of amino acid residues is given first, followed by the number of the residues deduced from the cDNA sequence; a dash (-) is inserted for proteins that were not predicted. The predicted localization as based on mainly STS data is given in the fourth column, followed by the exact localization of the genes (gene locus in bp as defined in the published sequences of chromosome 21, <http://hgp.gsc.riken.go.jp>, and chromosome 22, <http://www.sanger.ac.uk/cgi-bin/cwa/22cwa.pl>). The accession numbers of the genomic clone(s) covering the genes, identifiers of predicted transcripts (if available; dashes indicate nonpredicted genes), the number of predicted exons out of the number of identified exons (based on cDNA sequence), and the number of UniGene clusters that were hit with the respective cDNAs are given in columns 6–9.

Whereas 13 of the novel cDNAs map to chromosome 22, only two cDNAs map to chromosome 21. This could either be a reflection of the generally higher gene content of chromosome 22 (554 compared with the 225 predicted genes on chromosome 21) or be a result of the fact that the percentage of genes that had been known previously is higher for chromosome 21 (this chromosome had long been carefully investigated because of its clinical implications, e.g., in Down syndrome). A third explanation could be a correlation between chromosomal location and global expression levels of the individual genes, as has been proposed by Ewing and Green (2000), with genes mapping to chromosome 21 in general possibly being expressed at lower levels compared with genes located on chromosome 22.

By combining the genomic and cDNA data, the exact gene structures of all 15 cDNAs could be determined. Although all cDNAs were covered by UniGene clusters, only 8 of the 15 genes had been predicted from the genomic sequence. Most of these gene predictions were precise, identifying the majority or all exons. The number of amino acid residues varied in most cases only marginally from the number deduced from the cDNA sequence. However, one cDNA (DKFZp564B212) merged three predicted transcripts to only one gene and overlapped another gene (bK445C9.C22.3) predicted on the opposite strand. In total, seven genes had completely failed to be predicted, some of which encode rather large ORFs and consist of several exons.

The mapping information that is based on genomic sequence not only gives the exact localization of individual genes but also provides information on the context of these genes in view of neighboring

genes (e.g., DKFZp434B194 and DKFZp564B212 are only 13 kb apart) and the presence of probable additional gene copies. For example, the genes of cDNAs DKFZp434N035 and DKFZp434P211 appear to be present on chromosome 22 in 2 and 9 highly similar copies (>85% sequence identity on nucleotide level), respectively. DKFZp434P211 could indicate a cluster of highly similar POM121 related genes (Fig. 5), the first of which was described by Kawasaki et al. (1997). Two copies (2850458 and 2871777) seem to be ancient and inactive as they are incomplete, contain several frame shifts, and share only 89% and 87% sequence identity with the cDNA sequence in exon 1, respectively. The other copies are highly similar (>95% identity on nucleotide level). Further experiments will be necessary to investigate how many of the gene copies are expressed and to explain the presence of the stop codon at position 429 in three of the gene copies (and in the cDNA) but a sense codon in this position in four other gene copies, possibly leading to an extended protein product. EST evidence is available for transcripts of both types of genes (e.g., for copies 5055694 and 8220566).

DISCUSSION

The considerable fraction of genes that were not predicted in the analysis of the chromosome 21 and 22 sequences was somewhat surprising, as EST data and UniGene clusters (Table 3) were also available for these genes. Three of the genes that were not predicted even appear to be present in more than one copy on the same chromosome, namely, within 6 Mb on chromosome 22. But even if all genes could be identified via bioinformatic procedures, the alternative use of exons and promoters (alternative splicing) constitutes a problem that cannot currently be solved with knowledge of the genomic sequence alone. Consequently, only the availability of cDNA sequences enables us to define the precise protein coding parts of the genome and, in conjunction with the genomic counterpart, to also define the composition of exons in alternatively spliced transcripts of the same gene. Both the sequence and the chromosomal location of genes are important pieces of information supportive also in the process of defining and analyzing candidate disease genes.

Most of the genome has been unraveled as draft sequence, where sequence submissions of individual genomic clones are released in several contigs of varying length. These contigs are usually not ordered relative to one another. However, automated assembly and annotation tools like GoldenPath (<http://genome.ucsc.edu/goldenPath/hgTracks.html>) try to overcome this problem and prove to be extremely helpful for the mapping of cDNAs. The availability of cDNA sequences in turn immediately helps to identify the genes that are located on the respective

Table 3: Analysis of Gene Structures of cDNAs Mapping to Human Chromosomes 21 and 22									
Clone ID	Contig size (bp)	ORF size (aa)	Predicted chromosomal location	Chromosomal allocation relative to the published genomic sequences	Accession no. of genomic sequence entry	Annotation of predicted transcript	No. predicted/No. true exons	No. of UniGene clusters	
CHROMOSOME 21									
DKFZp434N0650	1095	186	21q11.2	2914115-2890884	HS21G104N		0/4	1	
DKFZp566A221	568	108/108	21q11.2	12684965-12674544	AL023107N		0/4	1	
CHROMOSOME 22									
DKFZp434B194	2876	838/837	22q11.2	10481146-10460626	Z95115	bk104819.C22.2	15/15	1	
DKFZp434F0116	2377	747/748	22q11.2	23317974-23333542	AL022312N		0/7	2	
DKFZp434H1150	3176	1445	22q11.2	26764656-26674550	AL049758		0/10	2	
DKFZp434N0331	1978	1160	22q11.2	4629050-4632539	AG007308N		0/8	2	
DKFZp434P211	5024	1428	22q11.2	53194302-5317804	AP000557N	AP000557.C22.1	0/8	1	
DKFZp564B212	1915	737/737	22q11.2	2699320-2701392	AC008132/8103		0/8	1	
DKFZp564E1978	1129	127/126	22q11.2	2850458-2715469	AC007326				
DKFZp564G1978	1662	422/423	22q11.2	4617790-4619361	AC007050				
DKFZp564K2478	1874	372	22q11.2	5055694-5042046	AC002472				
DKFZp564H2219	1971	476/476	22q11.2	5210385-5222231	AP000552				
DKFZp564I1978	1647	639/617	22q11.2	6560307-6558737	DB7013/DB700				
DKFZp564I1141	5071	588/617	22q11.2	8220566-8234258	AP000354				
DKFZp564I017121	2690	5/5	22q11.2	8627661-8626091	AP000356				
DKFZp564I017121	2690	5/5	22q11.2	10558798-10494469	Z95115	bk104819.C22.2	7/7	1	
DKFZp564I017121	2690	5/5	22q11.2	22474505-22491683	AL021707	bk145C9.C22.1			
DKFZp564I017121	2690	5/5	22q11.2	26895167-26845385	AL022476	bk145C9.C22.1			
DKFZp564I017121	2690	5/5	22q11.2	2498885-2525670	AC008079	dl508115.C22.1	5/5	1	
DKFZp564I017121	2690	5/5	22q11.2	12612231-2611609		dl508115.C22.1	10/11	2	
DKFZp564I017121	2690	5/5	22q11.2	12716855-2717477		dl508115.C22.1	0/11	2	
DKFZp564I017121	2690	5/5	22q11.2	15227647-5228269		dl508115.C22.1	0/11	2	
DKFZp564I017121	2690	5/5	22q11.2	295110375-29682479	Z953311/Z93784	bk145C9.C22.2	12/12	1	
DKFZp564I017121	2690	5/5	22q11.2	15147736-15179066	AG002073	AG002073.C22.1	16/15	2	
DKFZp564I017121	2690	5/5	22q11.2	25023143-25048485	AL055658	dl756C23.3	16/16	2	
DKFZp564I017121	2690	5/5	22q11.2	28333768-28351093	dl1033E15.1		0/2	1	

DKFZp434P211	1	MDSLWPGAG SHPPGVNTR LSPDLCPGKI VLRALKESG	-AGMPEQKD PRVQENPDQ RRVFETGDA	RPAPRPLADN GGLSPFVPGP	--GFLQTLLE	100
D87002		PS.....	
2699820	+	PS.....	
2850458	+R.....Q.....A.....S.....E.....D.....R.....O.....D.....	PS.....N.....SR.....ER.....	
2871777	-K.....AHNIQ LCPRLLSRED RVEGPPESGG	RDAAGQGQPS PRESSEKGSQ	HRGCTVCIAP VGQWRPLSLR AQANASARF	PCPEVESRIK
4617790	+S.....	PS.....	
5055694	+	PS.....	
5210585	+S.....	PS.....	
6560307	+	PS.....	
8220566	+	PS.....	
8627661	-	PS.....R.....	
DKFZp434P211	101	AQRSEIRYNG TSQTSWTSSC TNRNAISSFY SSTOGLPGIK	RRRGPASSHC QLTLSSTKV SEDRPAVSS	GHTQCEKAD IAPQQLTLE	NDSSTSEAGR	200
D87002		
2699820	+	
2850458	+VT.....R.....S.....M.....P.....PK.....	
2871777	-PHIPDLLEF	VEQTKCHREL LQVHGRLPVT	KEEGASLIPL	SAAPQVLKDS	BGPASGCLG
4617790	+D.....S.....	
5055694	+	
5210585	+	
6560307	+	
8220566	+P.....	
8627661	-	
DKFZp434P211	201	PSTHKPILLP RRRGEPLMLP PPLELOYVT	VEDLDREKA APORINSALQ	VEDKASDCR PSRPSHTLSS	LATGASGLPA VSKAPSDAQ	QETHKSGOCL
D87002		
2699820	+C.....	
2850458	+T.....I.....V.....	
2871777	-GASDAATSLT	AGVNGHCRPO	PGREGOTLVQ	KRTAGGQGHU	GLQTLTAPPH
4617790	+R.....H.....V.....A.....W.....C.....K.....
5055694	+P.....
5210585	+
6560307	+C.....
8220566	+
8627661	-W.....S.....	
DKFZp434P211	301	GLLDPLASAA GVPSTAPMSG	KMRPPGFLF	SSSDPLFATS	SDSQBSAQVT	SLIPAPFFAA
D87002	
2699820	+A.....
2850458	+K.....H.....
2871777	-
4617790	+VA.....T.....E.....E.....H.....
5055694	+
5210585	+
6560307	+
8220566	+
8627661	-
DKFZp434P211	401	ALHISGPOPO LQGVFRGQNG	RSQTSWTSSC	PK*		
D87002	R.....*	
2699820	+R.....RNAISSFY	RSTOGLPERK
2850458	-RRRGPASSHC
2871777	-QLNLSS*
4617790	+R.....RNAISSSC	SSTOGLPGRK
5055694	+R.....RNAISSFY	RSTOGLPERK
5210585	+R.....RNAISSFY	RSTOGLPERK
6560307	+R.....RNAISSFY	RSTOGLPERK
8220566	+R.....
8627661	-R.....

Figure 5 Multiple sequence alignment of cDNA DKFZp434P211 with POM121-related 1 (accession no. D87002) and sequences from chromosome 22 demonstrate the presence of a cluster of POM121-related genes. The individual genomic sequences were named after the start of the first exon relative to the cDNA: The open reading frame (ORF) was defined according to the predicted protein of the cDNA and of POM121-related 1. Genes located on the plus and minus strands of chromosome 22 are indicated with + and -, respectively. The cDNA sequence of DKFZp434P211 was taken as reference; identical residues in other sequences are indicated with a dot, residues deviating from the consensus are printed. Asterisks (*) indicate stop codons. The genomic sequences 2850458 and 2871777 are in italics because these copies deviate from the other copies by a premature stop or frame shifts and a large insertion, respectively, and are probably not expressed. In these two gene copies the initiator ATG is mutated. Dashes (-) were inserted by the software (CLUSTAL) to optimize the alignment.

genomic clones, to support the ordering of the draft sequence contigs, and to narrow down the regions where putative regulatory elements should reside. Thus, cDNA and genomic sequences are complementary and synergistically add information. The BLAST analysis of cDNAs and matching genomic sequences showed that only 32 cDNAs did not have corresponding genomic matches (not covered, NC in Table 2, column 5), which is the number expected because >91% of the genomic sequence are reported to be unraveled.

The chromosomal localization could be approximated for 449 cDNAs using the GoldenPath web browser; 21 BACs had not been mapped (NM). The accession numbers of these BACs are provided in column 5 of Table 2. The combination of genomic and cDNA sequence provides the gene structures with precise exon-intron boundaries and defined intron sequences.

Furthermore, it will become increasingly important to not only have the human genes identified but rather to characterize the precise functions of the en-

coded proteins and also the functions of those transcripts that are not translated. To this end, full-coding cDNA representations are indispensable tools, for example, for the subcloning of exactly defined ORFs into expression vectors. However, currently only ~11,000 nonredundant cDNA sequences have been deposited in public databases which are supposed to contain the complete protein coding ORF. An even lower number of these full-coding ORFs can be obtained as cDNA clones through commercial or noncommercial providers (e.g., ATCC, Genome Systems, Research Genetics, HGMP, Resource Center of the German Genome Project) and would thus be available for functional research.

Recently, the range of estimates given for the number of human genes has evolved to the lower end, because in two calculations only ~35,000 human genes have been predicted (Ewing and Green 2000; Roest Crolius et al. 2000). Our data would also hint at a lower than previously expected number, as we would estimate the number of genes currently represented in UniGene to be 50,000 at most. Still, the real number of human genes needs to be established by further cDNA and also by comparative genomic sequencing (e.g., of the mouse). If it should hold true, however, that the number of genes in human was indeed only about two-fold higher than the ~18,000 genes that have been predicted for *Caenorhabditis elegans* by The *C. elegans* Sequencing Consortium (1998) the question would arise as to where the difference in complexity between these two life forms originated. Because the sheer doubling of gene number would not be likely to account for all differences, the comprehensive analysis of gene and protein function(s) would become an even greater problem. This is because one solution to this apparent paradox could be the acquisition of multiple functions by many of the proteins expressed in human. This would add another order of complexity to the line starting with the genome and continuing through the transcriptome with alternative splicing, the proteome with post-translational modifications, and finally (?) to a 'functiome,' which would cover the acquisition of diverse functions by the same protein depending on its cellular and subcellular environment. Several examples of such multiple usages of proteins have already been described (Jeffery 1999).

In the set of 500 novel cDNAs described here, only about half of the deduced proteins could be functionally classified, while identification, for example, of a protein kinase does not provide information on substrates or pathways in which this protein is involved. Additionally, half of the predicted proteins remain without any hint as to their possible function. With this in mind, the establishment of a gene catalog which will eventually contain a nonredundant set of full-coding cDNA sequences and clones covering every

human gene, is prerequisite to carry out the experiments needed to precisely identify the protein function(s). This catalog should be the result of a global enterprise integrating the data and clones from as many projects and researchers as possible and could be an extension of already existing databases such as GeneCards (Rebhan et al. 1998) and RefSeq (Pruitt et al. 2000) with, for example, links to the clone providers mentioned above. In addition to the novel full-coding cDNA sequences and clones described here, we have identified over 1000 cDNAs which comprise full-coding representations of previously known genes. In combination, these cDNAs represent 2%–5% of all human genes and will thus be a substantial part of the catalog and be ideal tools to carry out functional analyses. Although the 500 novel cDNAs have been fully sequenced and can be directly used in functional analysis, the cDNAs representing known genes need further characterization because these are not fully sequenced. To this end, we amplify the ORFs from these cDNAs and verify the predicted size. These ORFs are then cloned into a bacterial expression vector which contains a N-terminal fusion with the GFP. As the Gateway system (Life Technologies) is employed in the cloning process, the ORFs can be shuttled into any expression vector (Simpson et al. 2000). Only intact reading frames (no PCR frame shifts, no introns, no frame shifts in the clone) lead to fluorescent colonies as the ORF extends uninterrupted into the GFP. The Gateway entry clones of the verified genes are also made available through the Resource Center.

To address the systematic functional analysis of the novel proteins, a large-scale project dealing with the subcellular localization and functional analysis of the proteins encoded by newly identified cDNAs reported here is underway (Simpson et al. 2000). Thus, the gene catalog in upcoming years will form the basis for the large-scale and comprehensive functional analysis of human genes and proteins, which is crucial to understand the basis of human life, disease, and death.

METHODS

Library Construction

SMART Libraries

The DKFZp564 (human fetal brain) and DKFZp566 (human fetal kidney) libraries were generated using the SMART kit (Clontech). PCR amplification of the cDNA was necessary to obtain enough cDNA for cloning. The first-strand primer did contain the KS sequence of the pBluescript vector (Stratagene) and any base but T (IUB code = V) in the 3'-terminal position of the primer [TCGAGGTCGACGGTATCGATAAG(T)₁₉V]. Amplification of the primary cDNA with Amplitaq (Perkin Elmer) and Pfu (Stratagene) DNA polymerases in a ratio of 19/1 (vol/vol) was carried out with primers that contained

uracil residues (3' primer: CAUCAUCAUCAUGAGGTCGACGGTATCGATAAG; 5' primer: CUACUACUACUATACGCTGCGAGAAGACGACAGAA) and that were compatible with the pAMP1 (Life Technologies) cloning sites for directional cloning. Prior to cloning, the cDNA was size fractionated on an agarose gel. Fragments >2 kb were excised and extracted from the gel using GELase (Epicentre). Cloning was done using uracil deglycosylase (UDG, Life Technologies) and chemically competent bacterial cells (XL-2 Blue, Stratagene).

Conventional Libraries

The DKFZp434 (human adult testis), DKFZp586 (human adult uterus), and DKFZp761 (human adult amygdala) libraries were generated using conventional approaches (Gubler and Hoffman 1983), employing a *NotI*-dT V primer for first-strand synthesis [GAGCGGCCGC(T)₁₉V]. After second-strand synthesis, *Sall* adapters were ligated to the blunt-ended cDNA. Then the cDNA was cut with *NotI* to generate *Sall*-*NotI*-compatible ends at the 5' and 3' ends of the cDNA, respectively, to allow directional cloning. The cDNAs were then size-selected on agarose gels in two dimensions and cloned into pSPORT1 pre-cut with *Sall* and *NotI* (Life Technologies).

Availability of cDNA Libraries and Clones

All libraries have been arrayed into 384-well microtiter plates and spotted on high-density nylon membranes. Each library consists of 27,000 clones or multiples thereof. High-density clone filters and individual clones are available through the Resource Center of the German Genome Project (<http://www.RZPD.de>; clone@pzd.de).

Selection of Clones for Sequencing

First, 5' ESTs were systematically generated from all clones of 384-well microtiter plates. The sequences were analyzed with BLASTN (Altschul et al. 1990) and BLASTX (Gish and States 1993) against EMBL, PIR, SWISSPROT, and TREMBL databases for the lack of identical (>95% identity over 50 bp) matches with known cDNAs, and for the presence of ORFs.

Clones with novel sequences were 3' end sequenced. These 3' ESTs were checked for the lack of matches with known genes in public databases, for repeat structures, and for the presence of polyadenylation signals. Clones matching the selection criteria were subjected to full-length sequencing.

Sequencing Methodology and Strategy

Sequencing was done preferentially using dye terminator chemistry (Applied Biosystems or Amersham) on ABI 377 automated DNA sequencers; one partner used EMBL prototype instruments (Wiemann et al. 1995) mainly with dye primer chemistry. Primer walking (Strauss et al. 1986) was the preferred sequencing strategy for the full-length sequencing of cDNAs. Design of walking primers was done preferentially using software (e.g., Schwager et al. 1995; Haas et al. 1998) that permitted the complete automation of this usually time-consuming process and thus helped in the parallel processing of large numbers of clones.

Bioinformatic Analysis

Every complete cDNA sequence was compared with the sequences in EMBL, EMBL-EST, EMBL-STS using BLASTN (Altschul et al. 1990). Searches against EMBL were done to determine whether the cDNAs were already known and to identify any genomic sequence information available that would cover the respective genes. Searches against EMBL-EST

were performed to analyze for the abundance of transcripts, to obtain information on a possible tissue specificity of expression, and to identify putative alternative splice forms or alternative use of polyadenylation signals. The annotations on the source tissue of the respective EST clones were parsed from the database entries to calculate the real ratio versus the expected ratio of expression according to the equation: (# hits tissue/total # hits)/(# ESTs tissue/total # ESTs). A gene that was transcribed at a constant level in many tissues would have a ratio of one. Significant higher or lower ratios would indicate increased or decreased levels of transcription in the tissue, respectively. To identify tissue-specific expression, the parameters were set to >4 ESTs matching the respective cDNA that needed to have been sequenced from a given tissue, and the cutoff for the ratio of overexpression was set to five. ESTs originating from pooled tissues or that were of unspecified origin were disregarded in this analysis. To obtain chromosomal mapping information, the sequences were aligned with the EMBL-STS database.

The potential protein-sequences were identified by a search for the longest ORF in each of the three forward frames with a minimum length of 90 codons. The deduced protein sequences were searched against the nonredundant protein data set of PIR, SWISSPROT, and TREMBL [BLASTP, using the SEG-filter by Wootton (1994)]. Any cDNAs without ORF >90 codons were analyzed with BLASTN against TREMBL to identify even shorter ORFs present.

BLASTX searches were performed against a nonredundant protein database comprising PIR, SWISSPROT, and TREMBL. The SEG-filter was used to screen for potential frame shifts in the coding sequences of the cDNAs and to identify cDNAs that were not fully spliced or were alternatively spliced. The protein sequence was then transferred to PEDANT (Frishman and Mewes 1997). PEDANT performed automated database searches: *psiBLAST* (Altschul et al. 1997), an iterated profile search procedure; *HMMER* (Sonnhammer et al. 1997), a Hidden Markov model software which uses statistical descriptions of a sequence family's consensus; and *BLIMPS* (Wallace and Henikoff 1992) for similarity searches against the BLOCKS (Henikoff et al. 2000) database. PROSITE protein sequence patterns were identified by ProSearch (Kolakowski et al. 1992). CLUSTAL-W (Thompson et al. 1994) was used for multiple sequence alignments of DNA and proteins. Transmembrane regions were identified by *ALOM2* (Klein et al. 1984), and signal peptides in secreted proteins by *SIGNALP* (Nielsen et al. 1997). SEG (Wootton and Federhen 1993) has been employed to detect low-complexity regions in protein sequences and *COILS* (Lupas et al. 1991) for the detection of coiled coils. For the functional classification of the cDNAs sequence, identities with *E*-values <10⁻³⁰ (BLASTN) and <10⁻¹⁰ (BLASTX) were accepted to be significant. The comprehensive bioinformatic data on all cDNAs analyzed by the Consortium are accessible at <http://www2.mips.biochem.mpg.de/proj/cDNA/index.html>. Mapping of the cDNAs to chromosomes was done first by BLAST analysis of the cDNA sequences against the human genomic sequence (NCBI-hgts database), followed by identifying the mapping position with help of the GoldenPath (Jim Kent, UCSC) browser (<http://genome.ucsc.edu/goldenPath/hgTracks.html>).

Availability of Clones and Further Information

All clones described here, and the other clones analyzed by the German cDNA Consortium, are available from the Resource Center of the German Genome Project (<http://www.RZPD.de>).

www.rzpd.de; clone@rzpd.de). The comprehensive bioinformatic data on all cDNAs analyzed by the Consortium are accessible at <http://www2.mips.biochem.mpg.de/proj/cDNA/index.html>. Additional information about the analysis of the described set of cDNAs is available at <http://www.dkfz-heidelberg.de/abt0840/GCC>. The full version of Table 2 can be obtained at this location in Excel, tab-delineated text, and pdf formats.

ACKNOWLEDGMENTS

We thank Christian Gruber, Oliver Heil, Lars Ebert, Daniel Bongartz, and Antje Krause for support with the bioinformatic analysis and presentation of the data, and Andreas Weller for encouraging discussions and support. This work has been supported by the Federal German Ministry of Education and Research (BMBF) through Projektträger DLR, within the framework of the German Human Genome Project (FKZ 01KW9705/07/10-16), and in part by the European Union (BIOMED 2 – BMH4-CT97-2284).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. 1998. New goals for the U.S. Human Genome Project: 1998-2003. *Science* 282: 682-689.
- Cross, S.H. and Bird, A.P. 1995. CpG islands and genes. *Curr. Opin. Genet. Dev.* 5: 309-314.
- Dunham, I., Shimizu, N., Roe, B.A., Chissole, S., Hunt, A.R., Collins, J.E., Bruskewich, R., Beare, D.M., Clamp, M., Smink, L.J., et al. 1999. The DNA sequence of human chromosome 22. *Nature* 402: 489-495.
- Ewing, B. and Green, P. 2000. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* 25: 232-234.
- Fields, C., Adams, M.D., White, O., and Venter, J.C. 1994. How many genes in the human genome? *Nat. Genet.* 7: 345-346.
- Frishman, D. and Mewes, H.-W. 1997. PEDANTIC genome analysis. *Trends Genet.* 13: 415-416.
- Gish, W. and States, D.J. 1993. Identification of protein coding regions by database similarity search. *Nat. Genet.* 3: 266-272.
- Gubler, U. and Hoffman, B.J. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263-269.
- Haas, S., Vingron, M., Poustka, A., and Wiemann, S. 1998. Primer design for large scale sequencing. *Nucleic Acids Res.* 26: 3006-3012.
- Hanke, J., Brett, D., Zastrow, I., Aydin, A., Delbruck, S., Lehmann, G., Luft, F., Reich, J., and Bork, P. 1999. Alternative splicing of human genes: More the rule than the exception? *Trends Genet.* 15: 389-390.
- Hanks, S.K., Quinn, A.M., and Hunter, T. 1988. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* 241: 42-52.
- Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., et al. 2000. The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature* 405: 311-319.
- Henikoff, J.G., Greene, E.A., Pietrokovski, S., and Henikoff, S. 2000. Increased coverage of protein families with the blocks database servers. *Nucleic Acids Res.* 28: 228-230.
- Jeffery, C.J. 1999. Moonlighting proteins. *Trends Biochem. Sci.* 24: 8-11.
- Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmelts, J.L., Wang, J., and Shimizu, N. 1997. One-megabase sequence analysis of the human immunoglobulin λ gene locus. *Genome Res.* 7: 250-261.
- Klein, P., Kanehisa, M., and Delisi, C. 1984. Prediction of protein function from sequence properties. Discriminant analysis of a data base. *Biochim. Biophys. Acta* 787: 221-226.
- Kolakowski, L.F., Jr., Leunissen, J.A., and Smith, J.E. 1992. ProSearch: Fast searching of protein sequences with regular expression patterns related to protein structure and function. *Biotechniques* 13: 919-921.
- Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* 234: 187-208.
- Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L., and Quackenbush, J. 2000. Gene index analysis of the human genome estimates approximately 120,000 genes. *Nat. Genet.* 25: 239-240.
- Lupas, A., Van Dyke, M., and Stock, J. 1991. Predicting coiled coils from protein sequences. *Science* 252: 1162-1164.
- Makalowski, W. and Boguski, M.S. 1998. Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci.* 95: 9407-9412.
- Nagase, T., Kikuno, R., Ishikawa, K., Hirose, M., and Ohara, O. 2000. Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.* 7: 143-150.
- Neer, E.J., Schmidt, C.J., Nambudripad, R., and Smith, T.F. 1994. The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371: 297-300.
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10: 1-6.
- Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayashi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K., and Tabata, S. 1994. Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (K1A0001-K1A00040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res.* 1: 47-56.
- Parraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T., and Kleit, R.E. 1988. Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* 241: 1489-1492.
- Pesole, G., Grillo, G., and Liuni, S. 1996. Databases of mRNA untranslated regions for metazoa. *Comput. Chem.* 20: 141-144.
- Pesole, G., Liuni, S., Grillo, G., Licciulli, F., Larizza, A., Makalowski, W., and Saccone, C. 2000. UTRdb and UTRsite: Specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* 28: 193-196.
- Pruitt, K.D., Katz, K.S., Sicotte, H., and Maglott, D.R. 2000. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet.* 16: 44-47.
- Rebhan, M., Chalfa-Caspi, V., Prilusky, J., and Lancet, D. 1998. GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 14: 656-664.
- Roest Crolius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fzames, C., Wincker, P., Brottier, P., Quetier, F., et al. 2000. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat. Genet.* 25: 235-238.
- Schuler, G.D. 1997. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J. Mol. Med.* 75: 694-698.
- Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., et al. 1996. A gene map of the human genome. *Science* 274: 540-546.
- Schwager, C., Wiemann, S., and Ansorge, W. 1995. GeneSkipper:

- Integrated software environment for DNA sequence assembly and alignment. *HUGO Genome Digest* 2: 8–9.
- Simpson, J., Wellenreuther, R., Poustka, A., Pepperkok, R., and Wiemann, S. 2000. Systematic subcellular localization of novel proteins identified by large scale cDNA sequencing. *EMBO Rep.* 1: 287–292.
- Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L., and Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci.* 91: 9228–9232.
- Sonnhammer, E.L., Eddy, S.R., and Durbin, R. 1997. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins* 28: 405–420.
- Strausberg, R.L., Feingold, E.A., Klausner, R.D., and Collins, F.S. 1999. The mammalian gene collection. *Science* 286: 455–457.
- Strauss, E.C., Kobori, J.A., Siu, G., and Hood, L.E. 1986. Specific-primer-directed DNA sequencing. *Anal. Biochem.* 154: 353–360.
- Struhl, K. 1989. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* 14: 137–140.
- Sutcliffe, J.G. and Milner, R.J. 1988. Alternative mRNA splicing: The Shaker gene. *Trends Genet.* 4: 297–299.
- Suzuki, Y., Ishihara, D., Sasaki, M., Nakagawa, H., Hata, H., Tsunoda, T., Watanabe, M., Komatsu, T., Ota, T., Isogai, T., et al. 2000. Statistical analysis of the 5' untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. *Genomics* 64: 286–297.
- The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* 282: 2012–2018.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673–4680.
- Wallace, J.C. and Henikoff, S. 1992. PATMAT: A searching and extraction program for sequence, pattern and block queries and databases. *Comput. Appl. Biosci.* 8: 249–254.
- Wiemann, S., Stegemann, J., Grothues, D., Bosch, A., Estivill, X., Schwager, C., Zimmermann, J., Voss, H., and Ansorge, W. 1995. Simultaneous on-line DNA sequencing on both strands with two fluorescent dyes. *Anal. Biochem.* 224: 117–121.
- Wootton, J.C. 1994. Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Comput. Chem.* 18: 269–285.
- Wootton, J.C. and Federhen, S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem.* 17: 149–163.
- Xu, N., Chen, C.Y., and Shyu, A.B. 1997. Modulation of the fate of cytoplasmic mRNA by AU-rich elements: Key sequence features controlling mRNA deadenylation and decay. *Mol. Cell. Biol.* 17: 4611–4621.

Received July 6, 2000; accepted in revised form December 29, 2000.